

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (uspto)**



PI Au-Young J;  
 XX  
 DR WPI; 2000-072605/06.  
 DR N-PSDB; AA56719.  
 XX  
 PT Proteins, polynucleotides, vectors, host cells and antibodies used to  
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,  
 PT neurological, gastrointestinal, developmental and cell proliferative  
 PT disorders -  
 PS  
 PS Claim 1; Page 128-130; 229pp; English.

CC AA56698 to AA56776 encode AAY57877 to AAY57955 which represent human  
 CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.  
 CC The transmembrane protein have immunospecific, antiproliferative and  
 CC neuroprotective activities. The human transmembrane proteins,  
 CC polynucleotides encoding them and other compositions and methods from  
 CC the present invention, can be used for the diagnosis, treatment or  
 CC prevention of immune, reproductive, smooth muscle, neurological, or  
 CC gastrointestinal, developmental and cell proliferative disorders. The  
 CC HTMPN's can be used to treat or prevent disorders associated with a  
 CC decreased expression or activity of HTMPN.  
 CC  
 SQ Sequence 688 AA;

Query Match 100.0%; Score 3639; DB 21; Length 688;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-234;  
 Matches 688; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSASGPGTRRLRLPYMGDLSTQSTQAOQOPANASTNPPETSNPKPRROT 60  
 DB 1 MSASGPGTRRLRLPYMGDLSTQSTQAOQOPANASTNPPETSNPKPRROT 60  
 QY 61 NOIQYLRLVLLKTLKHOFAFPOQVDAVKLMLPYKIKTKPMDSGTRKRLNNY 120  
 DB 61 NOIQYLRLVLLKTLKHOFAFPOQVDAVKLMLPYKIKTKPMDSGTRKRLNNY 120  
 QY 121 NMOECIODFNTMTNCTYINCKPDDIVLMAEALKFLQKINLPLEETIMIVQAKRG 180  
 DB 121 NMOECIODFNTMTNCTYINCKPDDIVLMAEALKFLQKINLPLEETIMIVQAKRG 180  
 QY 181 RGRKETGTAKPGVSTVNTQASTPOTQOPNPPVQATPPAPFVAVPDLIVQTFVMT 240  
 DB 181 RGRKETGTAKPGVSTVNTQASTPOTQOPNPPVQATPPAPFVAVPDLIVQTFVMT 240  
 QY 241 VVPQPLQTPRPVPOQPPAPAPQVQSHPIATPQPVKTKGKVKRKAADTTPTT 300  
 DB 241 VVPQPLQTPRPVPOQPPAPAPQVQSHPIATPQPVKTKGKVKRKAADTTPTT 300  
 QY 301 DPHEHPPSLPPEPKTKKIOGRSSRPVKPKKDVDSOHPAREKSKVSEQLKCCSGI 360  
 DB 301 DPHEHPPSLPPEPKTKKIOGRSSRPVKPKKDVDSOHPAREKSKVSEQLKCCSGI 360  
 QY 361 LKEMFAKKAHAAYAMPYKPDVAGLHLYCDIILKHPMDSTIKSLKLEAREYDAOEFGA 420  
 DB 361 LKEMFAKKAHAAYAMPYKPDVAGLHLYCDIILKHPMDSTIKSLKLEAREYDAOEFGA 420  
 QY 421 DVRLAFSNCITKYPDPHEVVAARKIQDVFEMFAKMPDEPEEPVAVSSPAVPPTKYV 480  
 DB 421 DVRLAFSNCITKYPDPHEVVAARKIQDVFEMFAKMPDEPEEPVAVSSPAVPPTKYV 480  
 QY 481 APPSSSDSSDSSDSSDSSDSEERAQRLAQLQELQKAVHQLAALSPQONKRRKKE 540  
 DB 481 APPSSSDSSDSSDSSDSSDSEERAQRLAQLQELQKAVHQLAALSPQONKRRKKE 540  
 QY 541 KDKKK 600  
 DB 541 KDKKK 600  
 QY 601 SEEDKCKPMSTYEEKROLSDINKLPGKIGRVVHIIQSREPSLKNSNDEIDETIK 660  
 DB 601 SEEDKCKPMSTYEEKROLSDINKLPGKIGRVVHIIQSREPSLKNSNDEIDETIK 660

QY 661 PSTLRELGALCHLLFAEEKETFKRLIM 688  
 DB 661 PSTLRELGALCHLLFAEEKETFKRLIM 688

RESULT 2  
 AAY07027  
 ID AAY07027 standard; Protein; 754 AA.  
 XX  
 AC AAY07027;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX

DE Breast cancer associated antigen precursor sequence.

XX Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.

OS Homo sapiens.

PN WO904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDWIG-) LUDWIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT Isolated using sera from cancer patients; used to develop products

PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure: Page 404-405; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.

XX Sequence 754 AA;

XX

Query Match 51.0%; Score 1857; DB 20; Length 754;  
 Best Local Similarity 57.7%; Pred. No. 2.9e-115;  
 Matches 390; Conservative 74; Mismatches 124; Indels 88; Gaps 19;

QY 44 NPPPETSNPKPKROTNOLOYLRLVLLKTLKHOFAFPOQVDAVKLMLPYKIKT 103  
 DB 13 NPPPETSNPKPKRGVNTNOLOYLRLVLLKTLKHOFAFPOQVDAVKLMLPYKIKT 72





```

XX AA07114;
AC 02-JUL-1999 (first entry)
DF WO904265 Seq ID No: 685.
DE Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX Homo sapiens.
XX MO9904265-A2.
XX 28-JAN-1999.
XX 15-JUL-1998; 98MO-US14679.
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896164.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfirsandshuh M, Sahlin U, Scanlan MJ, Stockert E;
PI Threlk O;
XX WPI: 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Disclosure: Page 728-730; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 947 AA:
XX
XX Query Match 41.3%; Score 1501.5; DB 20; Length 947;
XX Best Local Similarity 49.8%; Pred. No. 1,7e-91;
XX Matches 321; Conservative 78; Mismatches 154; Indels 91; Gaps 12;

```

```

OY 215 PPVQATPHPPFAVTPDLIVQTPVMTVPDPLOTPPPVPPQPPAPAPQVOSHPI 274
DB 154 IAVSSAKKESSESPATEKVKOOEIPSVFKTSI-----SPINVVQASV 197
OY 275 IATPQPVTKKGVKKRKAADTTPTTIDPIHEBPPSLPEPEKTKTKLGORRESSRPV--PK 332
DB 198 NSSQTAQVYTGVRKKAADTTTPAT--SAVKASSESP-----TTEKSVALLPPIKENMPK 251
OY 333 KDVPDSQQAPEBESKVSQOLCCSGILKEFARKHAAYAMPFYKPYDVEALGLDYCD 392
DB 252 NVLPDSQQAQYVVEYKVTLEQLRHCSSEILKEMIAKHFYAMPFYNPVDVNALGLHNYD 311
OY 393 ITKHPMDSTIKSLKAEKRYRDAQERGADEYRLAFNSCIYKYNPPDHEVYVAMAKKLDVFEK 452
DB 312 VYKNPMDLGTIKEMDNQYKDAYSPADYRILFMNICYKYNPPDHEVYVAMAKKLDVFEK 371
OY 453 RFKAMPDEPEE--PYVAASPAVPPTKVAAPSSSDSSSDSSSDSSDSDTDESEERQR 510
DB 372 HFSKIPPEVESMPICYITD-----TTEGTRENTNEASSSGNS--DSEDERYKR 422
OY 511 LAELQOLKAVHEOLAALSO--POQNKPKKKKKKKKKKK-----HKRKEVEENKKS 563
DB 423 LAKLQOLKAVHQQLQVLSQVFRKLKMKKKKKKKKKKKKKNNNNENNRKKEQMRKE 482
OY 564 KAKEPPPKTKKNNNSNSVSKKEPAPKSKPPPYESEEDKCKPMSTEEKROLSDIN 623
DB 483 KSKRNPQKKRKOQFIG-----LKSEEDNAPMNYDEKROLSLIN 523
OY 624 KLPGKLGKRVVHIIOSRPSLKNSPDEIEIDFETLKPTLREL 667
DB 524 KLPGKLGKRVVHIIOSRPSLKNSPDEIEIDFETLKASTREL 567

```

RESULT 5  
 ABB58985  
 ID ABB58985 standard; Protein: 1937 AA.  
 XX  
 AC ABB58985;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 3747.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 OS Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 DR WPI: 2001-656860/75.  
 DB N-PSDB; ABL03088.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Disclosure: SEQ ID NO 3747; 21pp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and



CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

**SQ Sequence 303 AA;**

Query Match	16.4%;	Score 598;	DB 22;	Length 303;
Best Local Similarity	45.4%;	Pred. No. 5.6e-32;		
Matches 132;	Conservative 37;	Mismatches 70;	Indels 52;	Gaps 9;

[illegible]

RESULT 7  
AAU16206  
ID AAU16206 standard; Protein; 235 AA.

AC AAU16206;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 1.159.

KM Human; immunosuppressive; antiarthritic; antirheumatic;  
KM cytostatic; caudant; vasotropic; cerebroprotective; nootropic;  
KM neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KM vulnereary; secreted protein; rheumatoid arthritis;  
KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KM cerebrovascular disorder; cerebral ischaemia; angiodysgenesis;  
KM nervous system disorder; Alzheimer's disease; infection; ocular  
KM corneal infection; wound healing; epithelial cell proliferation;  
KM skin ageing; food additive; preservative; antiproliferative.

**Homo sapiens.**

PN W0200155322-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01341.

PR	07-JUL-2000	2000S-0216680
PR	07-JUL-2000	2000S-0216647
PR	30-JUN-2000	2000S-0215135
PR	28-JUN-2000	2000S-0214866
PR	07-JUN-2000	2000S-0209467
PR	19-MAY-2000	2000S-0205513
PR	18-MAR-2000	2000S-0198123
PR	17-MAR-2000	2000S-0190076
PR	16-MAR-2000	2000S-0189354
PR	02-MAR-2000	2000S-0186750
PR	24-FEB-2000	2000S-0184664
PR	04-FEB-2000	2000S-0180628
PR	31-JAN-2000	2000S-0179665
PR	01-JAN-2000	2000S-0176589

PR	11-JUL-2000	2000US-0217486
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225313
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225667
PR	14-AUG-2000	2000US-0225687
PR	14-AUG-2000	2000US-0225570
PR	14-AUG-2000	2000US-0225577
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225759
PR	14-AUG-2000	2000US-0225758
PR	18-AUG-2000	2000US-0226679
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227170
PR	30-SEP-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229887
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229344
PR	05-SEP-2000	2000US-0229519
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	06-SEP-2000	2000US-0230436
PR	08-SEP-2000	2000US-0231142
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231145
PR	08-SEP-2000	2000US-0231141
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0233168
PR	14-SEP-2000	2000US-0233297
PR	14-SEP-2000	2000US-0233298
PR	14-SEP-2000	2000US-0233299
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0234584
PR	27-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235636
PR	29-SEP-2000	2000US-0236527
PR	29-SEP-2000	2000US-0236527
PR	29-SEP-2000	2000US-0236568
PR	29-SEP-2000	2000US-0236569
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240560
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809

PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PR XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI: 2001-488783/53.  
 DR N-PSDB; AAS26193.  
 DR XX  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 PS  
 PS Claim 11; SEQ ID NO 1159; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention.  
  
 Query Match 16.3%; Score 593; DB 22; Length 235;  
 Best Local Similarity 54.1%; Pred. No. 8.9e-32;  
 Matches 124; Conservative 21; Mismatches 56; Indels 28; Gaps 5;  
  
 QY 51 SNPKPKRQTNOLQYLRLVLTMTMKHOPAFPOQPVDAVKLNPDIYKIKTPMDMGTI 110  
 DB 1 SNPKPKRQTNOLQYLMONVYVTKMKHOPAFPOQPVDAIKNLDPYKIKTPMDMGTI 60  
 QY 111 KRLENNYYWNAQECIODFNTFTNCYIYNKPDIVIAEALFKLFLQKINLPTEETE 170  
 DB 61 KRLENNYYWNASQECODFNTFTNCYIYNKPDIVIAEALFKLFLQKINLPTEETE 120  
 QY 171 IMIVQAKGKRGKRETKAPGVSTVPNTQASTPQOTPPNPVPAOTAPHPAATVP 230  
 DB 121 ITTCSKG-----XKVGSRLLREPPAQVHSKWRPCPLSPQPPFACPPSPRPS 170  
 QY 231 DLIVQTPVMTVVP--QP-IGT-----PPVPPQGPQPPAPAPQVQSH 271  
 DB 171 SL-----PPLXQPSLQTSRRSQSPQLPPLLHPSPSPSLRH 209  
  
 RESULT 8  
 AAU16619  
 ID AAU16619 standard; Protein; 140 AA.  
 AC AAU16619;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human novel secreted protein, Seq ID 1572.  
 XX  
 KM Human; immunosuppressive; antiarthritic; antirheumatic;  
 KM cytoskeletal; cardiac; vasotropic; cerebroprotective; nootropic;  
 KM neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;  
 KM vulnerrary; secreted protein; rheumatoid arthritis;  
 KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KM cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KM corneal infection; wound healing; epithelial cell proliferation;  
 KM skin ageing; food additive; preservative; antiproliferative.  
 OS Homo sapiens.  
 XX  
 PN WO20015322 A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01341.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216680.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225457.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231443.  
 PR 08-SEP-2000; 2000US-0231444.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239335.  
 PR 13-OCT-2000; 2000US-0239337.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0250300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI  
 XX  
 PI  
 XX  
 DR  
 N-PSDB; AAS26606.  
 WPI: 2001-488783/53.  
 New nucleic acid molecules encoding 461 human secreted proteins for  
 diagnosis, preventing, treating or ameliorating medical conditions and  
 used as food additives or preservatives -  
 Claim 11; SEQ ID NO 1572; 980bp; English.

The invention relates to isolated nucleic acid molecules and their  
 encoded secreted proteins. The nucleic acids and proteins are used to  
 prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 rabbits, goats, horses, cats, dogs, chickens or sheep. They

CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Query Match 15.5%; Score 564; DB 22; Length 140;  
Best Local Similarity 75.4%; Pred. No. 4.1e-30;  
Matches 101; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QY 51 SNPKPKRQTNQQLYLRVLTLMKHQFAMPQOPVDVAVLNPYYKIKTPMDGTI 110  
DB 1 SNPKPKRQTNQQLYLRVLTLMKHQFAMPQOPVDVAVLNPYYKIKTPMDGTI 60  
QY 111 KKRLNNYVMAOECIDQFNMTNCTYNNKPGDIYVMAALRKLFLQKINELPTEETE 170  
DB 61 KKRLNNYVMAOECIDQFNMTNCTYNNKPGDIYVMAALRKLFLQKINELPTEETE 120  
QY 171 IMIYQAKGRGRK 184  
DB 121 IITPCSKGAKYGR 134

RESULT 9  
ABB68341.  
ID ABB68341 standard; Protein; 513 AA.

AC ABB68341;  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 31815.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

XX NO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PMD, Myers EW;  
PI WPI, 2001-656860/75.

XX DR N-PSDB; ABL12444.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX

PS Disclosure; SEQ ID NO 31815; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB157737-AB172072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 513 AA;

Query Match 11.8%; Score 430; DB 22; Length 513;  
Best Local Similarity 23.8%; Pred. No. 1.6e-20;  
Matches 151; Conservative 76; Mismatches 229; Indels 178; Gaps 21;

QY 26 MSTQQAQ-OPQANMASTNP-----PPETSNPKPKRQTNQQLYLRVLTLMKHQF 79  
DB 1 MNELOSNQNPPTNPPRPPYLVNGLVGPVPPRPPRPPRPPRPPRPPRPPRPPRPPRPP 59  
QY 80 AMPFQOPVDVAVLNPYYKIKTPMDGTIKRLNNYVMAOECIDQFNMTNCTYNN 139  
DB 60 SYHFRHPVDSVSLGPDVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 119  
QY 140 NKPGDDIYVMAALEKFLQKIN--ELPTEETIMYQAKGRGRKGTG----- 188  
DB 120 NLESPYVQAGKLMEAYNMEESIDISTE-----VELKPKSEKRRKATESIDQASTS 173  
QY 189 -AKPGVSTVPTQASFPQOTQOPNPBPVQA-----TPHPPAVNPDLIVQPP 238  
DB 174 FSAPRASNNYRQWLSSSSSMLCPRPMGVSFRPPSRNVRVGSILPSPFMDSLV-NPM 232  
QY 239 MTVPVPPPLQTPP-----PVPPQOPPPAPAPQVSHPPPIIATPPQVTKKGVKRRAD 293  
DB 233 QSMHPMSMNPPIFKNNMETNEADPPPS--EPISYRPLDSLAPLP----- 277  
QY 294 TTPPTTIDPIHEPPLP-PEPKTKLQGRRESSRPVPPKRDVDSOHPAP----- 345  
DB 278 -----SPMEPMTLPWPTP-----APVESPASSP-----PAPNPPIIICY 313  
QY 346 KSKVSEQLKCCSGILKMFKKHAAVAMP-----YKPYDVEALGLHYCDIK 395  
DB 314 KSLDRMTIKSHDHLMSVKKRKQYVAFNOADYMRYSQNP-----DYDHDRE 364  
QY 396 HPMDSTIKSLAREYDAQFADYRLMFSNCTYNNPDPHEVYVAMARKIQDFEMRFA 455  
DB 365 EKLDMKILQERLSDNDFESFPGFVSVYKRMQNALRCPEDEGLVYKSKTKNELFEKRLP 424  
QY 456 KMPDEPEEPVAVSSPAVPPTKYVAPPSSSSSSSSSSSSSSSDSDSDTDSSEERQRLAELQ 515  
DB 425 K-----YRELI 430  
QY 516 EQLKAVHEOLAALSOPOONKPKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 575  
DB 431 ATAKEKGRQLVA-----SRQDFRDSQNLILK-QENANKNKFGWPKPTIITKES 479  
QY 576 NNSNSNVSKKEPAPMKSKPP-----PTYESE 602  
DB 480 LFVOEQRVYKMEPDQVQEFSPGTEPLDQYVDS 513

RESULT 10  
AAG41763  
ID AAG41763 standard; Protein; 568 AA.

XX AAG41763;  
XX 18-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52002.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
PD  
XX 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 990S-0121825.  
PR 09-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 28-APR-1999; 990S-0130891.  
PR 30-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 04-MAY-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.

PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 02-JUL-1999; 990S-0142154.  
PR 06-JUL-1999; 990S-0142055.  
PR 08-JUL-1999; 990S-0142390.  
PR 09-JUL-1999; 990S-0142803.  
PR 12-JUL-1999; 990S-0142920.  
PR 13-JUL-1999; 990S-0142977.  
PR 14-JUL-1999; 990S-0143542.  
PR 15-JUL-1999; 990S-0143624.  
PR 16-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 19-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 22-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 26-JUL-1999; 990S-0145224.  
PR 27-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.





PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0148175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151338.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156599.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0162193.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 344; DB 21; Length 590;  
 Best Local Similarity 27.4%; Pred. No. 1e-14;  
 Matches 109; Conservative 49; Mismatches 156; Indels 84; Gaps 11;  
 QY 303 IHPPSLPEPKTKTGQRRESSRPYKPKKVP----DSQHPA-PEKSSKVSQKCC 357  
 DB 114 IPEVPAVRSAPLNNFTGKERNDLGPKKRKOKKNVSGIKRSNPGPDSPESEKTLIAGMLNFC 173

QY 358 SGIKEMFAKHAAYAMPFYKPDVVEALGLHDYCDIKNPDMSTIKSLEAREYDAQE 417  
CC the transcription activation characteristic of a signal dependent  
transcription factor.  
CC  
CC  
Db 174 SQILVTKLMKHK---WAWVFNTPDVVGIGLHDYHQVYKRPMDLGTVKLINTLKGFEVSPID 230  
QY 418 FGADVRLMFSCYKYNPPDHEVAMARKLDVFEKRFKAMDEDEEPVAVSSAIVPPT 477  
Db 231 FATVRLTFEDNMTYNPFGQDVFYMAKDLDFHFGMF-----NPAF 271  
QY 478 KVAAPSSDSSSDSSSDSTDDSEERARLAELOELK-----AVHOLAALSOPO 532  
Db 272 KKFPAQOLKLTGSSSRPPDPKPFQKQMNQNPVAMPKNGTEQISIAKKLSVAPQ 331  
QY 533 QNPKKK 592  
Db 332 PTLR-----POLVEPSRVQSPSPPPPPV-----IQPELPQP-- 363  
QY 593 SKPPPYESEEE-----DKCR-----SYEKKRLSDINKLPEEK 629  
Db 364 QPPPELEIEVADPDVSEVSKGRKGLPKPKAKDPNKRMTMEKSKLGNLADLPPEK 423  
QY 630 LGRVVAIIQSREPILKNSNPDEIEIDFETTLKPTLREL 667  
Db 424 LGQLQLIKRKHGHLAODG-DEIELDEAVDNETLMEL 460  
RESULT 12  
AAR79054  
ID AAR79054 standard; Protein; 2441 AA.  
AC AAR79054;  
XX  
XX 19-JAN-1996 (first entry)  
DE CREB binding protein.  
XX  
XX CREB binding protein; inhibition; promotion; activation; cAMP;  
KW cyclic adenosine monophosphate; reporter gene; reporter construct;  
KW transcription factor; identification; activation.  
XX  
OS Unknown.  
XX  
PN W09521940-A1.  
XX  
PD 17-AUG-1995.  
XX  
PF 01-FEB-1995; 95MO-US01325.  
XX  
PR 10-FEB-1994; 94US-0194468.  
XX  
PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
XX  
PI Montminy MR;  
XX  
DR WPI; 1995-293136/38.  
DR N-PSDB; AAQ99456.  
XX  
XX Identification of a compound having properties characteristic of CBP  
PT - which inhibits and/or promotes the activation of cAMP and mitogen  
PT responsive genes  
XX  
PS Claim 1; Page 32-39; 46pp; English.  
XX  
CC identification of a cpd. (I) which inhibits and/or promotes  
CC activation of cAMP and mitogen responsive genes comprises monitoring  
CC expression of a reporter in response to (I), relative to expression  
CC of reporter in the absence of (I), where exposure of (I) is  
CC performed in the presence of (1) a signal dependent transcription;  
CC (2) a polypeptide comprising at least amino acid residues 461-661 of  
CC this sequence (AAR79054); and (3) a reporter construct comprising a  
CC reporter gene under the control of a signal dependent transcription  
CC factor. The method can be used to identify a compound which has  
CC the binding and/or activation properties of CREB binding protein or

CC the transcription activation characteristic of a signal dependent  
transcription factor.  
CC  
CC  
SO Sequence 2441 AA;  
Query Match 8.9%; Score 323.5; DB 16; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 1.3e-12;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;  
QY 9 TRLRLNLPVNGGLFESQSTQAAQAPQANAASTNPP--PENSINPKPKRQNOLOYL 66  
Db 670 TRLHKGITLGN-----QPALPASQAPPVITPPAGVRRPN----- 704  
QY 67 LRVVLKTLMKHQFAMPFQPYDAVKLN--LPDYKIIITPMDMGITK-----KR 113  
Db 705 -----GRLPLVNRMQVSGMNSF-----NPSLSGVQLQAPKMGRAASP 745  
QY 114 LENNYVNAOECI-----ODFNTFTNCYINPKPDIVLMAEA-LEKLFLOKI 161  
Db 746 MNHSYQNMNSMASVPGMAISPSRMPOPRMNGTHA---NN-----IMAAQAPTONQFLPO- 795  
QY 162 NELPTEETEELMIVQAKGGRKRGKGTGA---KPGVS-----IVPNTQASTPPOQTP 211  
Db 796 NQFPS-SSGAMSVNSVSGQAPAAQAGVSGQEPGALPPLNMLAPQASQLPCEPVYQSP 854  
QY 212 -QPNPP-----VOATPBP-FPAVTPDLIVQTPVMTVPPPOPL---QTPPPVP----- 254  
Db 855 LHPTRPASTAAGMPSLQHPARPGMTF---PQPAAPTPSTPVSSGTPPTPTGSAVSA 910  
QY 255 -----POQPPPPAP-----APQPVOSHPEIT----- 275  
Db 911 AQOTSTPVQAAAOAOVTPPOQTPVQPPSVATPQSSQOQPPVHTQPGTFLSQAAASID 970  
QY 276 --AATPOVTKTKGKVKRKADTTP-----TITDPIHEPSPILPREKTKKLGQRSSS 325  
Db 971 NRVPPTSTVTAETSSQOQGVDPVLEKTEVQTDADPEPTEKGEPRSEMEEDLOGS 1030  
QY 326 RPVK-----PPKKVPPSQOHFAPE-----KSKVSEDLKCCSGILKEFAK- 368  
Db 1031 SQVKEETDTTRQKSEPMVEEKKEPVKVEAKEEENSNDPASSTSPQRRKIFRPEE 1090  
QY 369 -----NAAY-----AMPFYKPDVVEALGLHDYCDIKNPDMSTIKSLEAREYRD 414  
Db 1091 LROALMPTLEALYKQDPESELPFQPVDPQLGIDPDYIVKNPMDLSTIKRKLDTGQYE 1150  
QY 415 AQEFGADVRLMFSCYKYNPPDHEVAMARKLDVVEE 451  
Db 1151 PMQYVDVRLMFNNAWMLNKRKTSRYKFCSKLAEVFE 1187  
RESULT 13  
AAM40058  
ID AAM40058 standard; Protein; 2441 AA.  
AC AAM40058;  
XX  
XX 20-JUL-1998 (first entry)  
DE Cellular transcriptional factor CBP.  
XX  
XX Cellular transcriptional factor; CBP; P/CAF; human; transcription;  
KW histone acetyltransferase; HIV; infection; cancer; therapy;  
KW muscle differentiation.  
XX  
OS Homo sapiens.  
XX  
FH Key location/Qualifiers  
FT Misc-difference 377  
FT Misc-difference /note="encoded by CGN"  
XX  
PN W09803652-A2.  
XX  
PD 29-JAN-1998.





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:19:16 ; Search time 32 Seconds  
(without alignments)  
632.593 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639  
Sequence: 1 MSASGPGTRRLRLPYMGDG.....ALCHLFAEKEKFKLRKLM 688

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	330	9.1	65	1	US-08-227-536-5
2	330	9.1	65	5	PCT-US95-04682-5
3	323.5	8.9	2441	1	US-08-194-468-2
4	323.5	8.9	2441	3	US-08-961-739-2
5	323.5	8.9	2441	4	US-09-514-247A-8
6	314	8.6	65	1	US-08-227-536-6
7	314	8.6	65	5	PCT-US95-04682-6
8	308	8.5	2442	4	US-09-514-247A-10
9	305	8.4	2414	1	US-08-227-536-2
10	305	8.4	2414	5	PCT-US95-04682-2
11	209	5.7	331	6	5202236-37
12	209	5.7	865	4	US-09-281-766-19
13	207	5.7	941	4	US-07-757-022B-14
14	207	5.7	1022	4	US-07-757-022B-84
15	207	5.7	1038	4	US-07-757-022B-74
16	207	5.7	1049	4	US-07-757-022B-58
17	207	5.7	1140	4	US-07-757-022B-104
18	207	5.7	1270	4	US-07-757-022B-44
19	207	5.7	1311	4	US-07-757-022B-42
20	207	5.7	1313	4	US-07-757-022B-142
21	207	5.7	1314	4	US-07-757-022B-50
22	207	5.7	1320	4	US-07-757-022B-46
23	207	5.7	1330	4	US-07-757-022B-60
24	207	5.7	1354	4	US-07-757-022B-48
25	207	5.7	1361	4	US-07-757-022B-40
26	207	5.7	1363	4	US-07-757-022B-52
27	207	5.7	1404	4	US-07-757-022B-2

28	207	5.7	1404	4	US-07-757-022B-62	Sequence 62, Appl
29	206.5	5.7	1162	2	US-08-728-323A-2	Sequence 2, Appl
30	206.5	5.7	1162	4	US-09-298-568-2	Sequence 2, Appl
31	205	5.6	238	4	US-09-257-179-80	Sequence 80, Appl
32	204	5.6	1872	1	US-08-188-582-14	Sequence 14, Appl
33	204	5.6	1872	1	US-08-646-715-14	Sequence 14, Appl
34	204	5.6	1893	1	US-08-188-582-11	Sequence 11, Appl
35	204	5.6	1893	1	US-08-646-715-11	Sequence 11, Appl
36	202.5	5.5	334	6	5202236-3	Patent No. 5202236
37	202.5	5.5	1187	1	US-08-320-559-28	Sequence 28, Appl
38	200.5	5.5	1187	3	US-08-545-860D-28	Sequence 28, Appl
39	200.5	5.5	1187	5	PCT-US94-04496-28	Sequence 28, Appl
40	200.5	5.5	1210	1	US-08-320-559-26	Sequence 26, Appl
41	200.5	5.5	1210	3	US-08-545-860D-26	Sequence 26, Appl
42	200.5	5.5	1210	5	PCT-US94-04496-26	Sequence 26, Appl
43	199.5	5.5	214	1	US-08-217-327-4	Sequence 4, Appl
44	194.5	5.3	1185	4	US-09-041-886-23	Sequence 23, Appl
45	193.5	5.3	2972	4	US-09-579-181-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-227-536-5  
Sequence 5, Application US/08227536  
Patent No. 5658784  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPLICANT: Ewen, Mark  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,536  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DFCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYDROTHERICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-08-227-536-5

Query Match 9.1%: Score 330; DB 1; Length 65;

Best Local Similarity 87.7%: Pred. No. 6.7e-17; Mismatches 0; Gaps 0;

Matches 57; Conservative 4; Indels 0; Gaps 0;

```
DB 1 QRPMPROPVDAVKLGLPDYHKIKIPMDMGITKRLENNYWAASECQODFTMTTCY 60
QY 138 IYNKP 142
DB 61 IYNKP 65

RESULT 2
PCT-US95-04682-5
; Sequence 5, Application PC/TUS9504682
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,536
; FILING DATE: 14-April-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Heine, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-308Xq999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
PCT-US95-04682-5

Query Match 9.1%; Score 330; DB 5; Length 65;
Best Local Similarity 87.7%; Pred. No. 6, 7e-17;
Matches 57; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 78 QRPMPROPVDAVKLGLPDYHKIKIPMDMGITKRLENNYWAASECQODFTMTTCY 137
DB 1 QRPMPROPVDAVKLGLPDYHKIKIPMDMGITKRLENNYWAASECQODFTMTTCY 60
QY 138 IYNKP 142
DB 61 IYNKP 65

RESULT 3
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
```

```
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match 8.9%; Score 323.5; DB 1; Length 2441;
Best Local Similarity 23.4%; Pred. No. 1, 5e-14;
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLRNLPVAGDGLTSSQMTQAOAPANASTNPP--PETSNNPKRQTNOLQYL 66
DB 670 TRLRKGGIIGN-----QPALRASGAQPPYIPPAQSVRRPN----- 704
QY 67 LRVVLTKLKHOPMPPOQPVDAVKLN--LPDYKIKITPMDGTRK-----KR 113
DB 705 -----GRLPLPVNRMQVSGQNSP-----NPMISGVNQLPOAPKGPRAAP 745
QY 114 LENNYWNAOECI-----QDFNTMTTCYIYNKPGRDVIYMAA--LEKLFQKI 161
DB 746 MNHSVOMSNMSPGMAISPSRMPPOPPNMGTGA--NN-----TMAQAPTONOFLPO- 795
QY 162 NELPTEETIMIVQAKRGKRGKGTGA--KRGVS-----TVPNTQASTPOTOP 211
DB 796 NQFPS--SSGAMSVNSVGMQOPAAQAGVSGQEGGALPLPLNKLAPQAQSLRCPPTQSP 854
QY 212 -QNPMP-----VQATPHP--PPAVTPDLIVQTPVMTVVPPL--QTPPVV 254
DB 855 LHPTPPPASTAAGMPSLQHPGATP-----POPAAPTPQSTPVSQGPPTPTGSPVSA 910
QY 255 -----QOPQPPAP-----APQVQSPPLI----- 275
DB 911 AQTQSPVVOAAAQAQVTPQPPVQPPSVATPQSSQOQPTPHQPPPTPLSQAAASID 970
QY 276 --AATPQVTKKGVKRAKADTTP-----TTIDPHEPSLPEPPTTKLGGRESS 325
DB 971 NRPVTPSTVTSATSSQGPDPVPMLEMTYQTDIAEPPEPESKGEPSMEMEEDLOGS 1030
QY 326 RPYK-----PPKQVDPDSOHPAPE-----KSKVSEQLKCCSGILKEPKAKR- 368
DB 1031 SQVKEETDTTEQSEPMVEEKKPEVKVLAKEEENSNDTASQSTSPQPRKKIKRPE 1090
QY 369 -----HAAY-----AMPYTKPVDAVGLGHDYCDIHKHPMDNSTKSKLEAREYD 414
```



Db 1091 LRQALMPTLEALYKRODEBSLFRQPVDPQLGIDYEDYIVKPNMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYKFCSKLAIEVE 1187

RESULT 4  
US-08-961-739-2  
; Sequence 2, Application US/08961739A  
; Patent No. 6063583  
; GENERAL INFORMATION:  
; APPLICANT: Montminy, Marc R.  
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus  
; FILE REFERENCE: SALK1650-1  
; CURRENT APPLICATION NUMBER: US/08/961,739A  
; CURRENT FILING DATE: 1997-10-31  
; EARLIER APPLICATION NUMBER: US 194,468  
; EARLIER FILING DATE: 1994-02-10  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: Mus  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(2441)  
; OTHER INFORMATION: Xaa - Any Amino Acid  
US-08-961-739-2

Query Match 8.9%; Score 323.5; DB 3; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 1.5e-14;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLNLPLVMDGLETSMQSTTQAQAQOPANAASTNPP--PETSINPKPKROTNOLOYL 66  
Db 670 TRLKKGILGN-----QPALPASGAQPPVLPASVAPPN----- 704  
QY 67 LRVYLKILMKHQFAMPFOQPVDAVKLN--LPDYKIIKTPMDGKTIK-----KR 113  
Db 705 -----GRLPLPVNRMQVSGMNSF-----NPMSLGNVQLPQAPKGPRAASP 745  
QY 114 LNNYNNAAQECI-----QDFTMTNCTYNNKPPDDIVLMAEA--LEKLFLOKI 161  
Db 746 MNHSVQKNSMASVSGMALISPSRMPQPPMMGTHA--NN-----IMAAPTQNOQLPQ- 795  
QY 162 NELPTEETELMIYQAKGRGRKETGTA--KPGVS-----IVPNTTQASTPPQOTP 211  
Db 796 NQFPS--SSGASVSVSGMGQPAQAQVSGQGPALPPLMLAPQASQLPCPPVYQSP 854  
QY 212 -QPNPP-----VQATPRP--FPAVTPDLIVQTPVMTVPPQPL--QTPPPV- 254  
Db 855 LHPTRPPASTAAGMPSLQHPAPGKTP--POPAAPTQSTPVSQGTPTPTPGSVPSA 910  
QY 255 -----POQPPAP-----APQVOSHPII----- 275  
Db 911 AQTOSTPTVQAAQAQVTPQOTPVQPPSVATPQSSSQOQPPVHTOPGPTPLSQAASID 970  
QY 276 --AATPQPVKTKKGVKRRADTTT-----TTIDPIHEPPSLPPEKTKTKLQORSS 325  
Db 971 NRVTPTSVTSAETSSQOPGPDVPMLEKTEVQTDAPETTESKGEPRSSMEEDLQGS 1030  
QY 326 RPVK-----PRKQVPPDSQHPAPE-----KSKVSEOLKCCSGILKEFAK- 368  
Db 1031 SQVKEETDTTQKSEPEVEKEKPEVVEAKEEENSNDTASOSTSPQRRKIFKPEE 1090  
QY 369 -----HAAY-----AMPFYKPVDEVALGLHDYCDIHKPMDMSTIKSLEAREYD 414  
Db 1091 LRQALMPTLEALYKRODEBSLFRQPVDPQLGIDYEDYIVKPNMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
:: |||||:: || |||||

Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYKFCSKLAIEVE 1187

RESULT 5  
US-09-514-247A-8  
; Sequence 8, Application US/09514247A  
; Patent No. 6365361  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO. LTD.  
; APPLICANT: TANIGUCHI, Tomoyasu  
; APPLICANT: MIZUKAMI, Junko  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO  
; FILE REFERENCE: TANIGUCHI=6  
; CURRENT APPLICATION NUMBER: US/09/514,247A  
; CURRENT FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/03734  
; PRIOR FILING DATE: 1998-08-24  
; PRIOR APPLICATION NUMBER: JP231084/1997  
; PRIOR FILING DATE: 1997-08-27  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-514-247A-8

Query Match 8.9%; Score 323.5; DB 4; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 1.5e-14;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLNLPLVMDGLETSMQSTTQAQAQOPANAASTNPP--PETSINPKPKROTNOLOYL 66  
Db 670 TRLKKGILGN-----QPALPASGAQPPVLPASVAPPN----- 704  
QY 67 LRVYLKILMKHQFAMPFOQPVDAVKLN--LPDYKIIKTPMDGKTIK-----KR 113  
Db 705 -----GRLPLPVNRMQVSGMNSF-----NPMSLGNVQLPQAPKGPRAASP 745  
QY 114 LNNYNNAAQECI-----QDFTMTNCTYNNKPPDDIVLMAEA--LEKLFLOKI 161  
Db 746 MNHSVQKNSMASVSGMALISPSRMPQPPMMGTHA--NN-----IMAAPTQNOQLPQ- 795  
QY 162 NELPTEETELMIYQAKGRGRKETGTA--KPGVS-----IVPNTTQASTPPQOTP 211  
Db 796 NQFPS--SSGASVSVSGMGQPAQAQVSGQGPALPPLMLAPQASQLPCPPVYQSP 854  
QY 212 -QPNPP-----VQATPRP--FPAVTPDLIVQTPVMTVPPQPL--QTPPPV- 254  
Db 855 LHPTRPPASTAAGMPSLQHPAPGKTP--POPAAPTQSTPVSQGTPTPTPGSVPSA 910  
QY 255 -----POQPPAP-----APQVOSHPII----- 275  
Db 911 AQTOSTPTVQAAQAQVTPQOTPVQPPSVATPQSSSQOQPPVHTOPGPTPLSQAASID 970  
QY 276 --AATPQPVKTKKGVKRRADTTT-----TTIDPIHEPPSLPPEKTKTKLQORSS 325  
Db 971 NRVTPTSVTSAETSSQOPGPDVPMLEKTEVQTDAPETTESKGEPRSSMEEDLQGS 1030  
QY 326 RPVK-----PRKQVPPDSQHPAPE-----KSKVSEOLKCCSGILKEFAK- 368  
Db 1031 SQVKEETDTTQKSEPEVEKEKPEVVEAKEEENSNDTASOSTSPQRRKIFKPEE 1090  
QY 369 -----HAAY-----AMPFYKPVDEVALGLHDYCDIHKPMDMSTIKSLEAREYD 414  
Db 1091 LRQALMPTLEALYKRODEBSLFRQPVDPQLGIDYEDYIVKPNMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYKFCSKLAIEVE 1187  
RESULT 6

```
US-08-227-536-6  
Sequence 6, Application US/08227536  
Patent No. 5658784  
  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPlicant: Ewen, Maik  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,536  
FILING DATE: 14-APR-1994  
CLASSIFICATION: A36  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO.: 6:  
SEQUENCE CHARACTERISTICS:  
. LENGTH: 65 amino acids  
. TYPE: amino acid  
. TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEtical: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-08-227-536-6
```

Query Match            8.6%   Score 314; DB 1; Length 65;  
Best Local Similarity      84.6%; Pred. No. 9, 2e-16;

	Matches	55:	Conservative	4:	Mismatches	6;	Indels	0;	Gaps	0.
OY	371	AAYAWPYKPVVDDEALGLHDHDCDIIRHPMDMTIKSKLREAREYRDQAEFGADVRLMFNSCY								430
Dd	1	AAYAFWPKPVDASALSGLDHIITIKHPMDLSIVKKRMENRDYRDAGFAADVRLMFNSCY								60
OY	431	KYNPP	435							
db	61	KYNPP	65							

```
RESULT 7  
Sequence 6, Application PC/TU95504682  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
   APPLICATION NUMBER: PCT/US95/04682  
   FILING DATE:  
   CLASSIFICATION:  
   PRIOR APPLICATION DATA:  
   APPLICATION NUMBER: US 08/227,536  
   FILING DATE: 14-April-1994  
   CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
   NAME: Holliday C. Heine, Ph.D.  
   REGISTRATION NUMBER: 34,346  
   REFERENCE/DOCKET NUMBER: DCT-308xq999  
   TELECOMMUNICATION INFORMATION:  
     TELEPHONE: (617) 542-2290  
     TELEFAX: (617) 451-0313  
   INFORMATION FOR SEQ ID NO: 6:  
     SEQUENCE CHARACTERISTICS:  
       LENGTH: 65 amino acids  
       TYPE: amino acid  
       TOPOLOGY: linear  
       MOLECULE TYPE: peptide  
       HYPOTHETICAL: NO  
       ANTI-SENSE: NO  
       FRAGMENT TYPE: Internal

[illegible]

Db 783 MMAAPASQFLPQ-NQPPSSSGAMSV-----GNGOPRQGTGSOQVPGALPPLNMLG 837  
QY 197 PNTTQASTPQQTQP-OPNPPVQATPMPFPAVTEDLIYQTPVMTVPPQP-LQTPPEVP 254  
Db 838 PQASQLPCBPPTQSLHPTPPASTA-----AGMPSLQHTTP-PCMTPPQAPAPPTQSTP 891  
QY 255 -----PQPPPPAPAPQVQSHPIIA-----TPQ-----PVKTKKVKKRAKADTTTPTT 299  
Db 892 VSSSGQTPTPTGVSPTSQTQSTPTVQMAAQAVTPQPTQPTVQVPPVATPQSSQOQPT- 950  
QY 300 IDPIH-EPPSLPPEPKTKTLGQRESSRPV-----KPKKQVP-----DSOQ 340  
Db 951 --PVHAQPPGPTLSQAAASIDNRVPTBSSVASAETNSQOPQPDVLEMKTIETQAEDEP 1008  
QY 341 HPABEK-----SSKVSQOLKCCSGILKE----- 363  
Db 1009 DPGESKGPSEMEEDLOGASQVKEETDIAEQSEPMVEDEKPEVKEVEKEEESSEN 1068  
QY 364 -----MFAKKHAAY-----AMPFYKPVVEALGLHDYCDI 393  
Db 1069 GTASQSTSPQPRKKIRKPEELRQALMPTLEALYRQDESLPFRQPDVPLGIDYEDI 1128  
QY 394 IKHPMDSTIKSLKLEAREYRDOAEADVRLMFSNCYKPNPDHEVAMARKLDVFE 451  
Db 1129 VKNPMDSTIKRKLDGTQYQEPWQYVDDVWLMFNAMLYNKRKTSRYKFKCSKLAEFVE 1186

RESULT 9  
US-08-227-536-2  
Sequence 2, Application US/08227536  
Patent No. 5658784  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPLICANT: Ewen, Mark  
APPLICANT: Livingston, David  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,536  
FILING DATE: 14-Apr-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DFCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2414 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-227-536-2

Query Match 8.4%; Score 305; DB 1; Length 2414;  
Best Local Similarity 22.9%; Pred. No. 3e-13;  
Matches 133; Conservative 65; Mismatches 165; Indels 218; Gaps 24;

QY 9 TRL-LPVMGDLGLETQSMSTTQAOAPOPANASTNPPPEPS-----NPNKPK- 57  
Db 650 TRLQKNMLPNAAGVVPVSNBGPMMGQPP--GMTSNGPLDPSMIRGVSVPNQMPRTIT 707  
QY 58 -----ROTNOLQVLLRVLTKLKHQEFAMFQQPVDAVKINLDPY 97  
Db 708 POSGLOFGQMSMAOPPIVPRQTPLOH-----HGQADP----- 742  
QY 98 YKIITPMDKSTIKRLENNYWNQECIODFNTWTCYIYNKRGDDIVLMAEALKLF 157  
Db 743 -GALNPRGVPBMQOPNQGQFLPQ-----QFPFS----- 772  
QY 158 LQKINELPTEETEMIVQAKR--GGRKKTGTAKPGVSVVPTNTQASTPQQTQP- 213  
Db 773 -QGNM-----VTNIPPLAPSSQAPVSOQMSSSSCPVSPIIMPQSGSHIHCPQLQPA 826  
QY 214 -----NPPVQA-----TPHFPAVTPLIYQTPVMTVPPQLOTPPPVPPQ- 258  
Db 827 LHQNSPVPSPSTPTPHTP--PSIGAQPATYI-PAPVPTPAMPDGGQSQALHPP 882  
QY 259 -----PPAPAPQPVQSHPIIAATPQPVKTKGVRKADTTTPTIDPIHEPSPREP 313  
Db 883 RQTPPTPTQLPQVQVPLPAPASADQPO-----QPSQOSTASV-PTNAPLPLPQP 936  
QY 314 KTKLIGQ-----RRSSRPVKPPKK--DVPD----- 337  
Db 937 -ATPLSQPAVSIQGVSNPSTSTEVNSQAIKQPSQEVKMEKMEVDQPEPADTQPE 995  
QY 338 -----SOCHAPPEKSK--VSEDLK 355  
Db 996 DISKVEDCKWESTEERSTELKTEKEEDQPSISATQSPAPGSKKIRKPEELR 1055  
QY 356 CCSGILKEMFAKKHAAY-----AMPFYKPVVEALGLHDYCDIYKHPMDSTIKSLK 410  
Db 1056 -----QALMTLEALYRQDESLPFRQPDVPLGIDYEDIYKSPMDSTIKRKLDGT 1109  
QY 411 EYRDOEFGADVRLMFSNCYKPNPDHEVAMARKLDVFE 451  
Db 1110 QYEPWQYVDDVWLMFNAMLYNKRKTSRYKFKCSKLEFVE 1150

RESULT 10  
PCT-US95-04682-2  
Sequence 2, Application PC/TUS9504682  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04682  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,536  
FILING DATE: 14-April-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Holliday C. Heine, Ph.D.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: DFCI-308X999



Db 423 LITVGNMKEKRGKSKHFSLAESSNTGTSVQALPETDYST-----PQERDD 475  
QY 117 NYWNN-----AOEIODFTMTFCYIYNKPGDDIVLMAEALFKLQKINELPTEFEI 171  
Db 476 PYGKTKGTSMAPALSGIVALM-----LAANG-----LSATQIRSLMGSVRTPELSRV 527  
QY 172 -----MIVQAKRGKRGKREKGTAKPGVSTVNTQASTPPTQTPQTPQPNP 216  
Db 528 TWGAMPAPKRCIDALALVTPPEGRRPGNPP--SHPPPEASPE-----SSPPDQHHHPHP 581  
QY 217 ---VQATPHPRFAVTDLIQTVMTVVPVPPQIOTPPVPPVPPQPPAPAPP----- 267  
Db 582 RPNPEASPP--PEPS-----PPNMQHPRHPRPNRPNPEASPPSPPMNQ 624  
QY 268 -VQSHPRLIATPQPVTKKGVKKRKAADTTPTTIDPIHEPSSL-----PREPTTKLQ 320  
Db 625 HPHHP-----RNP-----PGASPPSPSPNMQHPRHPRPNP----- 661  
QY 321 RRESSRP--VKPRKDVDSQ-----QHAPREKSSVSEOLKCCSGILKEMFAKHA 370  
Db 662 --EASPPQSSPPEPQRFPSQWPHTPHFHYHPV----- 693  
QY 371 AYAMPF-----KPYDVEALGLHDYCDIKH--PMDMSTIKSL-----EAREYR 413  
Db 694 GYNLPYTYHQSLPYGPG--RDPCPCASHYPADDSPLGSYADDPSPQSPYPPSPSK 752  
QY 414 DAOEGADVRLMFNCYKNPDPHEVYVAMARKLDVFEFRPAKMD--EPEEPVAVSS 470  
Db 753 PSPPSGSPRV-----PSPHRHSRSLRPSAVESPPAPSPQSPHSPDTSFTK 803  
QY 471 PAVPPTKVVAP-----PSSSDSSSDSSSDSSTDDSEERA 508  
Db 804 PSTPPSPQDPQEGRRPESEED--DHKSLSDKSTSHSEGA 843

RESULT 13  
US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757, 022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luan  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-14

Query Match 5.7%; Score 207; DB 4; Length 941;  
Best Local Similarity 22.0%; Pred. No. 8.6e-07;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETELMIVQAKRGGRKE--TGAKPEVST--VQNTQASTPPTQOT--- 210  
Db 172 IKSAPTTKEKAPPTTKSAPTTKEKAPTTKEKAPTTKEKAPTTKEKAPTTKSAPT 231  
QY 211 ---POPNP--PVQATP--HPFPAVTDLIQTP-----VMTVPPQPLQTPP--- 251  
Db 222 TPKEAPPTTKKPAATTPKEKAPTTKEKAPTTKEKAPTTKEKAPTTKEKAPTPAKPK 291  
QY 252 ---PVPPQPPAPAPQVQSHPRPIAATPQPVYTKG--VKRAADTT----- 296  
Db 292 APTPKEKAPPTTKKPAATTPKEKAPTTKEKAPTTKEKAPTTKEKAPTTKEKAPTTK 351  
QY 297 ---PTTIDPIHEP--PSLPEPKTKRLGRRSSRPVKP--PKKDVDSQOHPAPEKSS 348  
Db 352 EPSPT---TKEPATPTKEKAPT-----TPKRAPPTKEKAPTTKEKAPTTK 399  
QY 349 KVSQOLKCCSGILKEMFAKHAAYAMPEYKPVDEALGLHDYCDIKHPMDSTIKSLTE 408  
Db 400 KPAPT-----APKEKAPTT-----KETAPTPKKLT 426  
QY 409 AREYRDAOEGADVRLMFNCYKNPDPHEVYVAMARKLDVFEFRPAKMDPEEP--- 464  
Db 427 PT-----TPKLAFTPEKAPTTPEEIAATTPPEPT-----PTPEKAPPT 469  
QY 465 --VAVSSPAVPPPT--KVAPPSSDSSSDSSSDSSTDDSEERAQRLAEQOLKA 520  
Db 470 PKAAPNTPEKAPPTTKKPAATTPKEKAPTTKEKAPTTKEKAPTTKEKAPTPPKP 529  
QY 521 VHEQLAALSOPOQNKPKKKEKKEKKEKKEKKEKKEEVENKSKAKEPPPKTKNNSSN 580  
Db 530 APKELAPPT---TKEPSTSDKPAFTPKGTATPTTPREKAPTTKEKAPTPPK---GT 582  
QY 581 SNVSKKEPAPKSKPPPTPE-----SEEDCKPKMSYE-----KRQ 617  
Db 583 APTLKEKAPPTPKKPAKELAPTTTKGTSTSDKPAFTPKETATPTTPREKAPTPPK 642  
QY 618 LSLDINKLPGKELGRVYHIISREPSLKNPNDE 651  
Db 643 PAPTTPPTPTTSEVSTPTTKETPTTIKSPDE 676

RESULT 14  
US-07-757-022B-84  
; Sequence 84, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143



```

QY 161 INELPTEETEMIVQAKGRGGRKE---TGTAKEGVST---VPNTQASTPPQOTOT--- 210
Db 237 IKSAPTTTPEKEBAPTTPKSAPTTTPKEBAPTTPKEBAPTTPKSAPTTTPKEBAPTTPKSAPT 296
QY 211 ----PQAMP--PQAMP--HPFPAVTPDLIVQTP-----VMTVVPQQLQTP--- 251
Db 297 TPKEBAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 356
QY 252 ----PVPPQPPAPAPAPQVOSHPIIAATPOPVYTKKG---VKRKADTTT----- 296
Db 357 APTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTTPKSAPTTPK 416
QY 297 ---PTTIDPIHP--PSLPEPEKTKLGQRRESSRPVKP---PKKQVPDSQOHPAPEKSS 348
Db 417 EPSPTT---TKEPAPTTPKEBAPT-----TPKKRPAPTTPKEBAPTTPKEBAPTTPK 464
QY 349 KVSQOLKCCSGILKEMFAKHAAYAMPYKPVVDVALGLHDYCDLIKHPMDSTIKSKLE 408
Db 465 KPAPT-----APKEBAPTTP-----KETAPTTPKKLT 491
QY 409 AREYRDAOEFGADYVIMFSNCKYKPPDHEVAVAMARKLDYFEMRFAMKMDPEEP--- 464
Db 492 PT-----TPKLIAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 534
QY 465 --VAVSSPAVPPT--KVAPSSSSSSSDSSSDSDSDSDSEERAQRLAELOQLKA 520
Db 535 PKAAPNTPEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 594
QY 521 VHEQLALISQOQONKPKKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 580
Db 595 APKEBAPTTP---TKEPSTSTSDKAPAPTTPKGTAPTTPKEBAPTTPKEBAPTTPK---GT 647
QY 581 SNVSKKEBAPKSKPPPTYE-----SEEDCKRPMSEYEB-----KRO 617
Db 648 APTLKEBAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 707
QY 618 LSLDINKLPGKELGRVYHI IOSREPSLKNSNPDE 651
Db 708 PAPTTPETPPPTTSEVSTPTTKETPTTIHKSPDE 741

```

Search completed: February 25, 2003, 13:23:25  
 Job time : 41 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compuser Ltd.

OK protein - protein search, using sw model

Run on: February 25, 2003, 13:20:47 : Search time 15 Seconds

(without alignments)  
1425.069 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639  
Sequence: 1 MSASGPGTRRLRLPVMGDG.....ALCHILPFAEKETFKRLKM 688

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*\n2: /cgn2\_6/ptodata/2/pubpaa/PTCT\_NEW\_PUB pep:\*\n3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*\n4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep:\*\n5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*\n6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*\n7: /cgn2\_6/ptodata/2/pubpaa/PTCTUS\_PUBCOMB pep:\*\n8: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*\n9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*\n10: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*\n11: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*\n12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*\n13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*\n14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	16.3	235	10	US-09-764-864-1159
2	564	15.5	140	10	US-09-764-864-1572
3	323.5	8.9	2441	12	US-10-109-886-8
4	308	8.5	2442	12	US-10-109-886-10
5	262.5	7.2	617	10	US-09-864-761-36182
6	220	6.0	128	10	US-09-764-864-1571
7	218.5	6.0	666	10	US-09-791-171-70
8	209	5.7	865	9	US-09-957-995A-19
9	207	5.7	941	12	US-10-124-557-14
10	207	5.7	945	8	US-08-965-272-2
11	207	5.7	945	12	US-10-121-882-2
12	207	5.7	1022	12	US-10-124-557-84
13	207	5.7	1038	12	US-10-124-557-74
14	207	5.7	1049	12	US-10-124-557-58
15	207	5.7	1140	12	US-10-124-557-104
16	207	5.7	1270	12	US-10-124-557-44
17	207	5.7	1311	12	US-10-124-557-42
18	207	5.7	1313	12	US-10-124-557-142
19	207	5.7	1314	12	US-10-124-557-50

20	207	5.7	1320	12	US-10-124-557-46	Sequence 46, Appl
21	207	5.7	1320	12	US-10-124-557-60	Sequence 60, Appl
22	207	5.7	1354	12	US-10-124-557-48	Sequence 48, Appl
23	207	5.7	1361	12	US-10-124-557-40	Sequence 40, Appl
24	207	5.7	1363	12	US-10-124-557-52	Sequence 52, Appl
25	207	5.7	1404	12	US-10-124-557-2	Sequence 2, Appl1
26	207	5.7	1404	12	US-10-124-557-62	Sequence 62, Appl
27	205.5	5.6	819	10	US-09-825-144-14	Sequence 14, Appl
28	205	5.6	238	10	US-09-729-835-80	Sequence 80, Appl
29	205	5.6	731	9	US-10-086-464-17	Sequence 17, Appl
30	203.5	5.6	555	10	US-09-764-864-1161	Sequence 1161, Ap
31	203.5	5.6	555	10	US-09-764-864-1574	Sequence 1574, Ap
32	201.5	5.5	5179	9	US-10-025-360-1068	Sequence 1068, Ap
33	201.5	5.5	5179	10	US-09-922-217-1068	Sequence 1068, Ap
34	201.5	5.5	5179	10	US-09-833-263-1068	Sequence 1068, Ap
35	197.5	5.4	580	10	US-09-764-864-1160	Sequence 1160, Ap
36	193	5.3	1125	9	US-09-974-298-114	Sequence 114, App
37	192.5	5.3	160	10	US-09-764-864-1165	Sequence 1165, Ap
38	192	5.3	449	10	US-09-764-864-1163	Sequence 1163, Ap
39	191.5	5.3	332	10	US-09-764-864-1573	Sequence 1573, Ap
40	191	5.2	2783	10	US-09-816-659A-14	Sequence 14, Appl
41	190.5	5.2	1192	10	US-09-758-140-6	Sequence 6, Appl1
42	190.5	5.2	1192	10	US-09-972-599A-6	Sequence 6, Appl1
43	189.5	5.2	1329	10	US-09-815-242-10112	Sequence 10112, A
44	187.5	5.2	1192	10	US-09-789-386-2	Sequence 2, Appl1
45	187.5	5.2	1192	10	US-09-893-348-23	Sequence 23, Appl

## ALIGNMENTS

RESULT 1  
US-09-764-864-1159  
Sequence 1159, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT23  
CURRENT APPLICATION NUMBER: US/09/764,864  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1159  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo.sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (129)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (215)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (221)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1159  
Query Match  
Best Local Similarity 54.1%; Pred. No. 2.3e-23;  
Matches 124; Conservative 21; Mismatches 56; Indels 28; Gaps 5;  
QY 51 SNPNKQOTNOLOVLLAVYKTKKHQFAPFQOPVDAVKIKNTLDYKRIITPMDKGTI 110  
DB 1 SNPSKPGKTNLOQYMNQVAVTKLKHQFAPFQOPVDAIKIKNTLDYKRIITPMDKGTI 60  
QY 111 KRLNNYVMAQECIODFNTFTNCYIYNNKGGDVIYMAEAEKLFLOKINELTEETE 170  
DB 61 KRLNNYVMAQECIODFNTFTNCYIYNNKGGDVIYMAEAEKLFLOKINELTEETE 120  
QY 171 IMIVQAKGKRGKRTGTAKKPGVSTVPNTQASTPPTQTPQPNPPVQATPPHPAVTP 230

```
Db 121 ITPCSKG-----XVGSRLREPRRAQVHSHKWPCCPLSDQRPFRACPPSPRRPS 170
Qy 231 DLIVGTPTVTVPP--QP-LQT-----PPVPPQPPPPAPAPQVOSH 271
Db 171 SL-----PPLVQPSLQTSRBSQSPQLPPHLLPHRSSPWSLLRH 209

RESULT 2
US-09-764-864-1572
; Sequence 1572, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1572
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1572

Query Match 15.5%; Score 564; DB 10; Length 140;
Best Local Similarity 75.4%; Pred. No. 3.9e-22;
Matches 101; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

Qy 51 SNBKRKQTNOLQYLRLVLTLMKHOEAMPPOQPDVAVKMLPDYKIKTKIPMDGTT 110
Db 1 SNBSKGRKTNQLYQNVVVKTKLMKHOEAMPPOPDVDAKMLPDYKIKTKIPMDGTTI 60

Qy 111 KRLNNYWNQECIODFTMTNCTYVKKPGDDIVYMAALEKLELQKINELPTEETE 170
Db 61 KRLNNYWNQECIODFTMTNCTYVKKPGDDIVYMAALEKLELQKINELPTEETE 120

Qy 171 IMIYQAKGRGRGRK 184
Db 121 ITPCSKAKVGSR 134

RESULT 3
US-10-109-886-8
; Sequence 8, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-10-109-886-8

Query Match 8.9%; Score 323.5; DB 12; Length 2441;
Best Local Similarity 23.4%; Pred. No. 6.1e-09;
```

```
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

Qy 9 TRLRLNLPVMDGIGLETSQMSITQAOAOPANMASTNPP--PETSNNPKRQTNOLQYL 66
Db 670 TRLHKQGIIGN-----QPALPSAGAPVPIPPAOSVRRPN----- 704

Qy 67 LRVVLTLMKHOPAMPPOQPDVAVKLN--LPDYKIKTKIPMDGTTK-----KR 113
Db 705 -----GRLPLPVNMQVSOGNSEF-----NMSIGNVQLPQAPGPPRAASP 745

Qy 114 LNNYWNQECI-----ODFTMTNCTYVKKPGDDIVYMAALEKLELQK 161
Db 746 MNSVQNSMASVPGMAISRSMPDPNNMGTA--NN-----IMAQAPYQNFLO- 795

Qy 162 NELPTEETELMIYQAKGRGRKGTGA--KPGVS-----TVPNTQASTPPQOTGP 211
Db 796 NQFPSS-SGSMASVSYGMQPPAAGVSGQGEAGALPNPLNLALQASQLPFPYVQTSF 854

Qy 212 -QNPVP-----VQATPH-PFVATPDLIVQTPVTVPPPPPL--QTPPVVP----- 254
Db 855 LHPTPPASTAAGMPSLQHPTAGMTP--PQPAAPQSTPVSSGQTPPTPGSVPSA 910

Qy 255 -----PQPPPPAP-----APQVQSHPII----- 275
Db 911 AQTQSTPTVQAAAQAOVTPQPPQVQPPSVATPQSSQOQPTPVHTQPTPLSQAAAASID 970

Qy 276 --AATPPVTKKGVKKRKAQDTTP-----TTIPNHPPLPPEPKTKLGGORRES 325
Db 971 NRPPTSTVTSATSSQPPQPDVPMLEKTEVQTDAAEFPTSKKEPSEMEEDLOGS 1030

Qy 326 RPYK----PPKQVDPSQHPAPE-----KSKVSSEQLKCCSGILKEMPAKK- 368
Db 1031 SQVKEETDTTEQSEMEVEEKKPEYKVEAKEEENSNDTASQSTSPQRRKIKFPEE 1090

Qy 369 -----HAAY-----AMPEKPYDVEALGHDVCDIIKHPMDSTKSKLEAREYD 414
Db 1091 LRQALMPTLEALYRQDESPFPQPPVDPOLGIPDIFDIYKKNMDLSTIKRLIDTQYDR 1150

Qy 415 AQEFGADVRLMFSNCYKYNPPDEHVVYAMARKLDQVE 451
Db 1151 PMQYVDVRLMFMNMLYNNKTSRYVKFCSKLALEVE 1187

RESULT 4
US-10-109-886-10
; Sequence 10, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 2442
; TYPE: PRT
; ORGANISM: human
US-10-109-886-10

Query Match 8.5%; Score 308; DB 12; Length 2442;
Best Local Similarity 26.1%; Pred. No. 3.6e-08;
Matches 109; Conservative 49; Mismatches 132; Indels 128; Gaps 19;
```

QY 148 LMAE-LEKLEFLQKINELPTEETIMIVQAKGRGKRETGTAK---PGVSTV----- 196  
Db 783 MMAQAPASQFLPQ-NQFPSSSGAMSV-----GMGQPPAQTGVSQGVGALPNPLNMG 837  
QY 197 PNTQASTPPTQOTQ-PQNPVPQATPHPPAVTPBDLVQTPVMTVPPQ-LQTPPPV 254  
Db 838 PQASQLPCCPVTQSPHLHTPPASTA-----AGMPSLQHTTP-PQMTPPQAPATQSPSTP 891  
QY 255 -----PQPPPPAPAPQPVQSHPIIA-----TPQ---PVTKKGVKKKADTTPTT 299  
Db 892 VSSSGQPTPTPTPGSVPSATQSTPTVQAAQAVTTPQPPQPPVQPPSATQSSQOQPT- 950  
QY 300 IDPIH-EPPSLPPEPKTKTKLQORRESSRPV-----KPKKDYV-----DSQ 340  
Db 951 --PVHAQPPGPTPLSQAAASINRVPTSSVASAETNSQOPDPVPLEMKTQEQADTEP 1008  
QY 341 HPAEEK-----SSKSEDLKCCSGILKE----- 363  
Db 1009 DPGESKGPRESMEEDLQASQVKEETDIAEQSEPMVEDEKKEPVKEVEKEEESSEN 1068  
QY 364 -----MFAKHAAY-----AMPFYKPVDEALGLHDYCDI 393  
Db 1069 CTASQSTSPSPQPKKIKRPEELRQALMPTLEALYRQDEPSLPPQPPVDPQLGTPDIFDI 1128  
QY 394 IKHPMDSTIKSLKLEAREYRDQEFQADVRLMFSNCRKYNPDDEHVAMARKLQDVEE 451  
Db 1129 VKNMDLSTIKRKIDTQCYQEPQYVDVWLMFNAMLYNKRKTSRVYKFCSKLAEVFE 1186  
RESULT 5  
US-09-864-761-36182  
; Sequence 36182, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 36182  
;; LENGTH: 617  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005529.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.94  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6  
;; OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUATE 2.00e-33  
;; OTHER INFORMATION: EST\_HUMAN HIT: BF445754.1, EVALUATE 3.00e-22  
US-09-864-761-36182  
Query Match 7.2%; Score 262.5; DB 10; Length 617;  
Best Local Similarity 21.9%; Pred. No. 1.8e-06;  
Matches 126; Conservative 96; Mismatches 268; Indels 85; Gaps 20;  
QY 138 IYKPPGDDIVIMAELEKFLQKINELPTEETIMIVQAKGRGKRETGTAKPGVSTVP 197  
Db 43 VKESEKETEIVIEQOTER---TQVTEVEETEEKEKEKEEGEGEEBAEG----- 94  
QY 198 NTQASTPPTQOTQPP---NPPVQATPHPPAVTPBDLVQTPVMTVPPQ-LQTPPPV 253  
Db 95 -EEETKSPPAEASPEKEAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 153  
QY 254 PPOPPAPAPAPQVOSHPIIAATPQPVKT--KKGVKKADTTPTTI-DPIHEPSLP 310  
Db 154 EAK-SPEPAKSPKEEAKSPAEVKSPEKAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 212  
QY 311 PEPTTKLQORRESSRPVKKKQVDSQOHAPAEKSSVSEDLKCCSGILKEPAKKA 370  
Db 213 AEAKSPVKEEAKSPAEVKSPEKAKSPTEKEAKSPKEAKSP-----KEEA 257  
QY 371 AYAMPFYKPVDEALGLHDYCDIIRKHPMDSTIKSLKLEAREYRDQAEFGADVRLMFSNCY 430  
Db 258 KSPKAKSPVAEA-----KSP---EKASPVAA-EAKSPKAKSPVKEAKSP 303  
QY 431 KYNPDDEHVAMARKLQDVEFEMREKMPDEPEEV-----VAVSPAVPPTKVVAPP 483  
Db 304 KAKSPVKEAKSPKAKSPVKEE-AKTEPAKSPVKEEAKSPKAKSPKAKSPKAKSPKAKSP 362  
QY 484 SSSSSSSSSSDSDSDSDSEERAQRLAELOEQLKAVHEDLALISPOQKPKKKEKD- 542  
Db 363 AKTAPAKEEARSADRF--PEKAKSPVKEEYKSPKAKSPKLAOKAKPAKEKIPKKEEYKS 419  
QY 543 --KKEKKKEKHKREVEENKSKAKPEPPPKTKNNNSNVSNSVSKKEPAPKSPPTPYE 600  
Db 420 PVKEEKPQEVKVEPPKKAEEKAPATPKTEEKD-----SKKEAPKKEKAPKVE 472  
QY 601 SEEE--DKCPMSYEERQSLDINKL--GEEKLGRVYHIIQGREPSLK---NSNPE 651  
Db 473 EKKEPAVEKPKRESKVEAKKEAEADKKVPTPEKEAPKAVEKEDAKPREKTEVAKKREDD 532  
QY 652 IEIDPEILKPESTLELGLALCHLLAEKEKTFKLK 686  
Db 533 AKAK-EPSKPAEKKEA-----APEKKDTKEEK 558  
RESULT 6  
US-09-764-864-1571  
; Sequence 1571, Application US/09764864

```

; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1571
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1571

```

```

Query Match      6.0%; Score 220; DB 10; Length 128;
Best Local Similarity 37.9%; Pred. No. 5.5e-05;
Matches 44; Conservative 20; Mismatches 50; Indels 2; Gaps 1;

QY 48 PETSNNKRRKRONOQYLRLVYLKTLKHKOPAPQOPDPAVKLMDPIYKLIKTPMDM 107
DB 9 PSKREXSKRPDQJSTLSTLSKLSQVKSQSMPEPVK--RTAPGYEYIRPMDL 66
QY 108 GTRKRLNNYNNNAOECIODENTMTFCYIYKPGDDIYLMALAEKLEKLINE 163
DB 67 KTMSELRKRYYSKRLPMADLRVLTNCKEYNPSEYIKCANILEKFFPSKIKE 122

```

```

RESULT 7
US-09-791-171-70
; Sequence 70, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WEDJINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: DERIVED FROM M. TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 70
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-70

```

```

Query Match      6.0%; Score 218.5; DB 10; Length 666;
Best Local Similarity 21.8%; Pred. No. 0.00031;
Matches 98; Conservative 33; Mismatches 156; Indels 16; Gaps 15;

QY 190 KRGVSTVPNTQ-----ASTPQOTQPOPNNPVQATPPHPAVPMDLIVQTPVMTVP 243
DB 44 KPGQGTPTPTSDLSRFRFASAPPPPPPPPPPP-----TWPDIAGHPSPSEPAASKPTP 100
QY 244 PQLQTPPPVPPQOPPPAP--APQVQSHP-----PIAATPQPVKTKKGVKRAKDTTT 296
DB 101 PMPIAGPEPAPKPPPTPPPIAGPEPAPKPPPTPPPIAGPATPTESQIAPPPPTPT 160
QY 297 PPTID-----PIHPPSLPPPKTKTKLGQRESSR 326
DB 161 PTGAPQPPSPAPHPVSHGHQRTAPAPPMKMMIGEPAPRSR-----SAS 210
QY 327 PVKPPKVDVDSQHPAPEKSKVSQOLCCSGILKEMFAKKAHAAYAMPYKVDYALG 386
DB 211 PAEPPTP-----PAPQHSRR-----ARRGHRRTDTERNVGKVAATG 246
QY 387 LHDYCDIHKHPMDSTIKSLAREYRDAQDFGADVRLMFSNCKYKNPDQHEVVAARKL 446
DB 247 -----PSTQARLRAEASGAQ-----LABQTE----- 268
QY 447 ODVFENRFAKMPDEPEPVVAVSSPAVPPTKVVAPPPSSSDSSSDSSDSTDSSEB 506
DB 269 -----PSPAPLQGPSYLAPTRPAPTE---PPSPSPQNRNG-----R 304
QY 507 RAQRLAEQQLKAVHEQIALAISQPOONKPKKKEKKKKKKKKRKEVEENKSKAK 566
DB 305 RAER-----RVHPDLAAQHAQAQPSITAATGRRRRKRAADLDLATORSLRPAK 355
QY 567 EPPPKTKKNNSSNVSKKEPAPMKRPP 596
DB 356 GKPVKKVPKQK-----PKATKPP 373

```

```

RESULT 8
US-09-957-995A-19
; Sequence 19, Application US/09957995A
; Patent No. US20020165373A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Patricia A.
; APPLICANT: Barr, Bradd C.
; APPLICANT: Anderson, Mark L.
; APPLICANT: Sverlow, Karen W.
; TITLE OF INVENTION: Recombinant Neospora Antigens and Their
; Uses
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/957,995A
; FILING DATE: 15-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,858
; FILING DATE: 21-MAR-1994
; APPLICATION NUMBER: US 08/327,516
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: US 08/645,951
; FILING DATE: 10-MAY-1996

```

APPLICATION NUMBER: US 09/281,766  
FILING DATE: 30-MAR-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-082512US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 865 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-957-995A-19

Query Match 5.7%; Score 209; DB 9; Length 865;  
Best Local Similarity 21.8%; Pred. No. 0.0012;  
Matches 127; Conservative 54; Mismatches 214; Indels 188; Gaps 29;

QY 12 RNLVPMGGLTSMSTQAOAQRAPN-----ASTNPPRPSNPNKPKRQTNOLQY 65  
DB 363 RGADIMHSWGSDBASELRQAIQOTAOQGIHHAAGNSGRDVTPTYPALSTAVEG 422  
QY 66 LIRVVLKTIAMKHO-----FAMFQOPVDVAKLNP--DYKKIKITPMDMTIKRLN 116  
DB 423 LITVGNMKEKQROSKHFLAESSNVTGKSQIALPGTDIYSTI-----PVGERDD 475  
QY 117 NYTN-----AOEIQDENFTNCTIYNKRGDDIVLMAELKFLQKINELPTEETEI 171  
DB 476 PYGMKTGTSMAPALSGIVALM-----LAANPG-----LSATQIRSLIMOSVNTPELSTRV 527  
QY 172 -----MIYQAGRGGRKGTGAKPGVSTVPTTQASNPQOTQOPR 216  
DB 528 TWGAMPAKRLDALVLPREGRRGNP--SHRPEASPP--SSPRDQNHPR 581  
QY 217 ---PVQATPRPRAVTPDLIVQTVYVPPRQPLQTPRPVPPRQPRPARAPQ----- 267  
DB 582 RPNPEASR-DEPS-----PRNQHPRHPRPRPRPRPRPRPRPR 624  
QY 268 -VQSHPRITATPQPVTKKKGKAKADTTPTTIDRIHERPSL-----PREPKTKKLGQ 320  
DB 625 HPRHPR-----PCASPRSSPPRMOHPRHPRPRPRPRPRPRPR 661  
QY 321 RRESSRP-VKRPKKQVDPDSO-----QHAPREKSSKVSQOLKCSGILKEMPAKKA 370  
DB 662 --EASPPQSSPPRQPRPSQMPHTRHFNHPR-----EAREYR 413  
QY 371 AYAMPFY---KPYDVEALGLHYCDILKH--PMDMTIKSL-----EAREYR 413  
DB 694 GYNLPYFYHQSPRLPGYRG-RDPCQASHNRYRADDSPLGVARPSPQSVPRPRPSK 752  
QY 414 DAOEGADVRLMFSNCYINPRDHEVYVAMARKLDVFEKRAKMD---EEEPYVAVSS 470  
DB 753 PSPEGSPPRV-----PSPHNPRSKLPSAVERSPRPSQSPRPSPTSPK 803  
QY 471 PAVPPPTKVAVP-----PSSSDSSSDSSSDSTDDSEERA 508  
DB 804 PSTPPSPSQDPEGRREPSSEED---DHKSLDKSTSHSSEGA 843

RESULT 9  
US-10-124-557-14  
Sequence 14, Application US/10124557  
Patent No. US20020137894A1  
GENERAL INFORMATION:  
APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-APR-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cegitt, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-124-557-14

Query Match 5.7%; Score 207; DB 12; Length 941;  
Best Local Similarity 22.0%; Pred. No. 0.0016;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETIMTVQAKGRGRKE---TGAKPGVST---VPTTQASTPPTOTOT--- 210  
DB 172 IKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 231  
QY 211 ---PQNP--PVQATP-HPRPAPTPLIVQTP-----VNTVPPQQLQRP--- 251  
DB 232 TPKEPAPTTPKPAVTPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 291  
QY 252 ---PVPPQPPRPAPOVOSHPPITATPQPVTKKG---VKKADTTT----- 296  
DB 292 APTTPKEPAPTTPKPAVTPTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPK 351  
QY 297 ---PTTIDPIHPR-PSLPPPEPTTKLGGRRSSRAVVP---PKADVDSOOHRAPEKSS 348  
DB 352 EPSPTT---TKEPAPTTPKEPAPT-----TPKKPAPTTPKEPAPTTPKEPAPTTPK 399  
QY 349 KVSQOLKCCSGILKEMPAKKAHAAYAMPYKPYDVEALGLHYCDILKHPRMDMTIKSL 408  
DB 400 KPAPT-----APKEPAPT-----KEPAPTTPKLT 426  
QY 409 AREYDAQEGADVRLMFSNCYINPRDHEVYVAMARKLDVFEKRAKMPDEPEP----- 464  
DB 427 PT-----TPKLAPTTPKEPAPTTPPELAPTTPEPT-----PTTPEPAPT 469  
QY 465 --VAVSSPAVPPPT--KVAVPSSSDSSSDSSSDSTDDSEERAOLAELODLKA 520





```

1 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
2 NUMBER OF SEQUENCES: 143
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: Genetics Institute, Inc.
5 STREET: 87 CambridgePark Drive
6 CITY: Cambridge
7 STATE: Massachusetts
8 COUNTRY: U.S.A.
9 ZIP: 02140
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16
17 CURRENT APPLICATION NUMBER:
18 APPLICATION NUMBER: US/10/124,557
19 FILING DATE: 16-Apr-2002
20 CLASSIFICATION: <Unknown>
21
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 07/643,502
24 FILING DATE: 18-JAN-1991
25 APPLICATION NUMBER: US 07/546,114
26 FILING DATE: 29-JUN-1990
27 APPLICATION NUMBER: US 07/457,196
28 FILING DATE: 29-DEC-1989
29 APPLICATION NUMBER: US 07/390,901
30 FILING DATE: 08-AUG-1989
31 ATTORNEY/AGENT INFORMATION:
32 NAME: Geert, Luann
33 REGISTRATION NUMBER: 31,822
34 REFERENCE/DOCKET NUMBER: GI 5190
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (617)876-1170
37 TELEFAX: (617)876-5851
38
39 INFORMATION FOR SEQ ID NO: 74:
40 .
41 .
42 .
43 .
44 .
45 .
46 .
47 .
48 .
49 .
50 .
51 .
52 .
53 .
54 .
55 .
56 .
57 .
58 .
59 .
60 .
61 .
62 .
63 .
64 .
65 .
66 .
67 .
68 .
69 .
70 .
71 .
72 .
73 .
74 .
75 .
76 .
77 .
78 .
79 .
80 .
81 .
82 .
83 .
84 .
85 .
86 .
87 .
88 .
89 .
90 .
91 .
92 .
93 .
94 .
95 .
96 .
97 .
98 .
99 .
100 .
101 .
102 .
103 .
104 .
105 .
106 .
107 .
108 .
109 .
110 .
111 .
112 .
113 .
114 .
115 .
116 .
117 .
118 .
119 .
120 .
121 .
122 .
123 .
124 .
125 .
126 .
127 .
128 .
129 .
130 .
131 .
132 .
133 .
134 .
135 .
136 .
137 .
138 .
139 .
140 .
141 .
142 .
143 .
144 .
145 .
146 .
147 .
148 .
149 .
150 .
151 .
152 .
153 .
154 .
155 .
156 .
157 .
158 .
159 .
160 .
161 .
162 .
163 .
164 .
165 .
166 .
167 .
168 .
169 .
170 .
171 .
172 .
173 .
174 .
175 .
176 .
177 .
178 .
179 .
180 .
181 .
182 .
183 .
184 .
185 .
186 .
187 .
188 .
189 .
190 .
191 .
192 .
193 .
194 .
195 .
196 .
197 .
198 .
199 .
200 .
201 .
202 .
203 .
204 .
205 .
206 .
207 .
208 .
209 .
210 .
211 .
212 .
213 .
214 .
215 .
216 .
217 .
218 .
219 .
220 .
221 .
222 .
223 .
224 .
225 .
226 .
227 .
228 .
229 .
230 .
231 .
232 .
233 .
234 .
235 .
236 .
237 .
238 .
239 .
240 .
241 .
242 .
243 .
244 .
245 .
246 .
247 .
248 .
249 .
250 .
251 .
252 .
253 .
254 .
255 .
256 .
257 .
258 .
259 .
260 .
261 .
262 .
263 .
264 .
265 .
266 .
267 .
268 .
269 .
270 .
271 .
272 .
273 .
274 .
275 .
276 .
277 .
278 .
279 .
280 .
281 .
282 .
283 .
284 .
285 .
286 .
287 .
288 .
289 .
290 .
291 .
292 .
293 .
294 .
295 .
296 .
297 .
298 .
299 .
300 .
301 .
302 .
303 .
304 .
305 .
306 .
307 .
308 .
309 .
310 .
311 .
312 .
313 .
314 .
315 .
316 .
317 .
318 .
319 .
320 .
321 .
322 .
323 .
324 .
325 .
326 .
327 .
328 .
329 .
330 .
331 .
332 .
333 .
334 .
335 .
336 .
337 .
338 .
339 .
340 .
341 .
342 .
343 .
344 .
345 .
346 .
347 .
348 .
349 .
350 .
351 .
352 .
353 .
354 .
355 .
356 .
357 .
358 .
359 .
360 .
361 .
362 .
363 .
364 .
365 .
366 .
367 .
368 .
369 .
370 .
371 .
372 .
373 .
374 .
375 .
376 .
377 .
378 .
379 .
380 .
381 .
382 .
383 .
384 .
385 .
386 .
387 .
388 .
389 .
390 .
391 .
392 .
393 .
394 .
395 .
396 .
397 .
398 .
399 .
400 .
401 .
402 .
403 .
404 .
405 .
406 .
407 .
408 .
409 .
410 .
411 .
412 .
413 .
414 .
415 .
416 .
417 .
418 .
419 .
420 .
421 .
422 .
423 .
424 .
425 .
426 .
427 .
428 .
429 .
430 .
431 .
432 .
433 .
434 .
435 .
436 .
437 .
438 .
439 .
440 .
441 .
442 .
443 .
444 .
445 .
446 .
447 .
448 .
449 .
450 .
451 .
452 .
453 .
454 .
455 .
456 .
457 .
458 .
459 .
460 .
461 .
462 .
463 .
464 .
465 .
466 .
467 .
468 .
469 .
470 .
471 .
472 .
473 .
474 .
475 .
476 .
477 .
478 .
479 .
480 .
481 .
482 .
483 .
484 .
485 .
486 .
487 .
488 .
489 .
490 .
491 .
492 .
493 .
494 .
495 .
496 .
497 .
498 .
499 .
500 .
501 .
502 .
503 .
504 .
505 .
506 .
507 .
508 .
509 .
510 .
511 .
512 .
513 .
514 .
515 .
516 .
517 .
518 .
519 .
520 .
521 .
522 .
523 .
524 .
525 .
526 .
527 .
528 .
529 .
530 .
531 .
532 .
533 .
534 .
535 .
536 .
537 .
538 .
539 .
540 .
541 .
542 .
543 .
544 .
545 .
546 .
547 .
548 .
549 .
550 .
551 .
552 .
553 .
554 .
555 .
556 .
557 .
558 .
559 .
560 .
561 .
562 .
563 .
564 .
565 .
566 .
567 .
568 .
569 .
570 .
571 .
572 .
573 .
574 .
575 .
576 .
577 .
578 .
579 .
580 .
581 .
582 .
583 .
584 .
585 .
586 .
587 .
588 .
589 .
590 .
591 .
592 .
593 .
594 .
595 .
596 .
597 .
598 .
599 .
600 .
601 .
602 .
603 .
604 .
605 .
606 .
607 .
608 .
609 .
610 .
611 .
612 .
613 .
614 .
615 .
616 .
617 .
618 .
619 .
620 .
621 .
622 .
623 .
624 .
625 .
626 .
627 .
628 .
629 .
630 .
631 .
632 .
633 .
634 .
635 .
636 .
637 .
638 .
639 .
640 .
641 .
642 .
643 .
644 .
645 .
646 .
647 .
648 .
649 .
650 .
651 .
652 .
653 .
654 .
655 .
656 .
657 .
658 .
659 .
660 .
661 .
662 .
663 .
664 .
665 .
666 .
667 .
668 .
669 .
670 .
671 .
672 .
673 .
674 .
675 .
676 .
677 .
678 .
679 .
680 .
681 .
682 .
683 .
684 .
685 .
686 .
687 .
688 .
689 .
690 .
691 .
692 .
693 .
694 .
695 .
696 .
697 .
698 .
699 .
700 .
701 .
702 .
703 .
704 .
705 .
706 .
707 .
708 .
709 .
710 .
711 .
712 .
713 .
714 .
715 .
716 .
717 .
718 .
719 .
720 .
721 .
722 .
723 .
724 .
725 .
726 .
727 .
728 .
729 .
730 .
731 .
732 .
733 .
734 .
735 .
736 .
737 .
738 .
739 .
740 .
741 .
742 .
743 .
744 .
745 .
746 .
747 .
748 .
749 .
750 .
751 .
752 .
753 .
754 .
755 .
756 .
757 .
758 .
759 .
760 .
761 .
762 .
763 .
764 .
765 .
766 .
767 .
768 .
769 .
770 .
771 .
772 .
773 .
774 .
775 .
776 .
77
```

```

Db 535 PKAAPNPPEKPAPTPKPEAPPTPKPEAPPTPKETAAPTTKGTAPTTLKEKPAPTPKPK 594
Oy 521 VHEQJLAALSQPOPNPKKKKKKKKKKKKKRKEVEENKSKAKAPPKTKKNNSSN 580
Db 505 ADKELAPPT---TKBPTSTTSDDKPAPTTTPKGTAPTTTPEKPAPTPKPKPK---GT 647
Oy 581 SNVSKKEPAPMKSKPPPTYE-----SEEDCKCKPMSEYEE-----KRO 617
Db 648 APTTLKEKPAPTPKPKAPKELAPTTTKGTSTTSDDKPAPTTPEKPAPTTTPKPKAPTTPKK 707
Oy 618 LSLDINKLPGKELGRVHIIQSRPESLKNSSNDE 651
Db 708 PAPPTTPEPTTSEVSTPTTKEPTTIHKSPDE 741

RESULT 14
US-10-124-557-58
; Sequence 58, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 5.7%; Score 207; DB 12; Length 1049;
Best Local Similarity 22.0%; Pred. No. 0.0018:
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23.
Oy 161 INELPTEEFIMIVAKGGRGRKE---TGAKPGVST---VPMTQASTPQGTOT--- 210

```



Db 280 IKSAPITPKKBPATTTKSAPTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKSAPT 339  
QY 211 ----PQNP--PVQATP--HFFAVTPDLIVQTP-----VMVVPQPLQTP-- 251  
Db 340 TPKEBPATTTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 399  
QY 252 ----PVPPQOPPPAPAPQVQSHPIIAATPOPVKTKG--VKRADTTT----- 296  
Db 400 APPTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTK 459  
QY 297 ----PTIDPIHEP--PSLPPREKTKLGORRESSPVKP--PKKDVPDSQOHAPPEKSS 348  
Db 460 EPSPTT--TKEPATTTKEKBPATTT-----TPKKBPATTTKEKBPATTTKEKBPATTTK 507  
QY 349 KVSQOLKCCSGILKEMFAKHAAYAMPYKRVDEVALGLHDYCDLIHMDMSTIKSKLE 408  
Db 508 KPAFT-----APKEPATTT-----KETAPTTPKLT 534  
QY 409 AREYDAQEGADYRLMFSNCKYNPDPHEVAMARKLDVFEKRFKMPDEPEP-- 464  
Db 535 PT-----TPKLAFTTPEKBPATTTPELAPTTPEEP--PTTPEBPATTT 577  
QY 465 --VVAVSSPAVPPPT--KVVAAPSSSDSSSDSSSDSSSDSSSEERAQRLAELQOLKA 520  
Db 578 PKAAPPTPEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 637  
QY 521 VHEDLALISPOQKPKKKEKKKKKKKKKKKKKEVEENKSKAKBPPTTKKNNSSN 580  
Db 638 APKEPATTT--TKEPSTISDKAPPTTKEKBPATTTKEKBPATTTKEKBPATTTK--GT 690  
QY 581 SNVSKKEPAPKSKPPPYE-----SEEDKCKPMSYEE-----KRQ 617  
Db 691 APPTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTK 750  
QY 618 LSLDINKLPGEKLGKVVHIIOSREPSLKSNDPE 651  
Db 751 PAPTTPPTTSEVSTPTTKEPTTIHKSNDPE 784

RESULT 15  
US-10-124-557-104  
; Sequence 104, Application US/10124557  
; Patent No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesener, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserf, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1140 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 104:  
US-10-124-557-104

Query Match 5.7%; Score 207; DB 12; Length 1140;  
Best Local Similarity 22.0%; Pred. No 0.0019;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETELIMIVQAGRGGRKE--TGAKPGVST--VPNTTOASTPQOT-- 210  
Db 371 IKSAPITPKKBPATTTKSAPTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 430  
QY 211 ----PQNP--PVQATP--HFFAVTPDLIVQTP-----VMVVPQPLQTP-- 251  
Db 431 TPKEBPATTTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 490  
QY 252 ----PVPPQOPPPAPAPQVQSHPIIAATPOPVKTKG--VKRADTTT----- 296  
Db 491 APPTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTK 550  
QY 297 ----PTIDPIHEP--PSLPPREKTKLGORRESSPVKP--PKKDVPDSQOHAPPEKSS 348  
Db 551 EPSPTT--TKEPATTTKEKBPATTT-----TPKKBPATTTKEKBPATTTKEKBPATTTK 598  
QY 349 KVSQOLKCCSGILKEMFAKHAAYAMPYKRVDEVALGLHDYCDLIHMDMSTIKSKLE 408  
Db 599 KPAFT-----APKEPATTT-----KETAPTTPKLT 625  
QY 409 AREYDAQEGADYRLMFSNCKYNPDPHEVAMARKLDVFEKRFKMPDEPEP-- 464  
Db 626 PT-----TPKLAFTTPEKBPATTTPELAPTTPEEP--PTTPEBPATTT 668  
QY 465 --VVAVSSPAVPPPT--KVVAAPSSSDSSSDSSSDSSSDSSSEERAQRLAELQOLKA 520  
Db 669 PKAAPPTPEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 728  
QY 521 VHEDLALISPOQKPKKKEKKKKKKKKKKKKKEVEENKSKAKBPPTTKKNNSSN 580  
Db 729 APKEPATTT--TKEPSTISDKAPPTTKEKBPATTTKEKBPATTTKEKBPATTTK 781  
QY 581 SNVSKKEPAPKSKPPPYE-----SEEDKCKPMSYEE-----KRQ 617  
Db 782 APPTKBPATTTKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKEKBPATTTKBPATTTK 841  
QY 618 LSLDINKLPGEKLGKVVHIIOSREPSLKSNDPE 651  
Db 842 PAPTTPPTTSEVSTPTTKEPTTIHKSNDPE 875

Search completed: February 25, 2003, 13:23:44  
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:18:31 ; Search time 47 Seconds

(without alignments)  
1407.244 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639  
Sequence: 1 MSASGPGTRRLRLPVVWGDS.....ALCHLLPAEERETFKLRKLM 688

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869.5	51.4	733	2 T28145	RING3 kinase - chi
2	1857	51.0	754	2 A56619	female sterile hom
3	1339.5	36.8	2038	2 A43742	female sterile hom
4	608.5	16.7	1087	2 T22847	hypothetical prote
5	600.5	16.5	1250	2 T22845	hypothetical prote
6	570.5	15.7	686	2 S55955	bromodomain protei
7	561	15.4	578	2 T40984	transcription fact
8	495.5	13.6	638	2 S67605	hypothetical prote
9	368.5	10.1	361	2 T42517	bromodomain protei
10	339.5	9.3	461	2 D96757	hypothetical prote
11	339	9.3	678	2 T49984	bromodomain protei
12	323.5	8.9	766	2 A86198	hypothetical prote
13	323.5	8.9	2441	2 S39161	CREB-binding prote
14	318	8.7	440	2 H86312	F2H15.2 protein -
15	310	8.5	2440	2 S39162	transcription coac
16	305	8.4	2414	2 A54277	transcription adap
17	297.5	8.2	374	2 T33328	hypothetical prote
18	285	7.8	1188	2 S49915	extensin-like prot
19	282	7.7	369	2 T46098	hypothetical prote
20	267.5	7.4	703	2 T48600	kinase-like protei
21	267.5	7.4	1020	1 QPHUH	neurofilament trip
22	265.5	7.3	400	2 T00472	probable RING3 pro
23	265	7.3	2027	2 S60123	hypothetical prote
24	265	7.3	2056	2 G88564	protein R10E1.1 f
25	264	7.3	971	2 T19431	hypothetical prote
26	260	7.1	1072	1 A37221	neurofilament trip
27	259.5	7.1	1274	2 T16251	hypothetical prote
28	258	7.1	854	2 S02003	neurofilament trip
29	254.5	7.0	7962	2 I38346	elastic titin - hu

30	251.5	6.9	1087	1 QFM5H	neurofilament trip
31	248.5	6.8	1633	2 JCS056	polybromo 1 - chic
32	243	6.7	606	2 A43427	neurofilament trip
33	240	6.6	856	2 T16543	hypothetical prote
34	238	6.5	1879	2 T19481	hypothetical prote
35	237	6.5	2649	2 T51023	hypothetical prote
36	234.5	6.4	228	2 S53504	hypothetical prote
37	233	6.4	769	2 E96613	extensin-like prot
38	233	6.4	847	2 F96531	hypothetical prote
39	232	6.4	1474	2 T20488	hypothetical prote
40	231	6.3	699	2 T38073	hypothetical prote
41	231	6.3	1211	2 T42230	nucleolar phosphop
42	225	6.2	6642	2 T29757	Af4 protein - mous
43	224	6.2	832	2 S71788	protein UNC-89 - C
44	223.5	6.1	369	2 S20500	P/CAF protein - hu
45	223.5	6.1	1006	2 T42731	hydroxyproline-ric atrophin-1 related

#### ALIGNMENTS

##### RESULT 1

T28145

RING3 kinase - chicken

C:Species: Gallus gallus (chicken)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T28145

R:Milne, S.; Kaufman, J.; Beck, S.

Submitted to the EMBL Data Library, May 1998

A:Description: DNA sequencing and analysis of the chicken major histocompatibility co

A:Reference number: 220475

A:Accession: T28145

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-733 <MT>

A:Cross-references: EMBL:AL023516; NID:e1292539; PID:e1292549; PIDN:CAA18965.1

A:Experimental source: clone cbl2

A:Gene: RING3

A:Map position: 16

A:Introns: 64/3; 110/3; 158/1; 227/3; 351/3; 394/3; 479/3; 546/2; 650/1; 691/1

C:Superfamily: unassigned bromodomain proteins; bromodomain homology

F:52-109/Domain: bromodomain homology <BR01>

F:323-380/Domain: bromodomain homology <BR02>

Query Match

Best local similarity 51.4%; Score 1869.5; DB 2; Length 733;

Matches 391; Conservative 69; Mismatches 142; Indels 69; Gaps 14;

QY	26	MSITQAQAPQPPANASTNPPESTSNPKPKQTNLQYLRLVLTLMKHQFAPRQ	85
DB	1	MASVVALDTLPQ-----ANPPEVSNPKKPRQVYNQQLYLVKYMALMKHQFAPRQ	54
QY	86	PVDAAKMLPYYKIKTPMDGKIKRLNNYVWNAOECIQDFMTFTNCYINRPGDD	145
DB	55	PVDAAKMLPYYKIKTPMDGKIKRLNNYVWNAOECIQDFMTFTNCYINRPGDD	114
QY	146	IYMAEALKEFLKINLPTRETIMVYQKGRGRKGTAPGVSTPNTQASTP	205
DB	115	IYMAQTEKIFLQVAAQMPDEDEIVVPAK-----NSHKKGASRAAALLAGLTA	165
QY	206	PQOTPPQNPVPQATPPFPAVTPDLIVQTPVMTVPPQPLQTPPVPPOQPPAPAP	265
DB	166	AO-----QVAVSVSVST-AVYTPSPDIATTVI-----PSPVSAFL	205
QY	266	QPVOSHPPIAATP--QPVYTKKGVKKRADTTPPTIDPI-----HPPSLAPPKTK	317
DB	206	KSLHSTAVLVAPAPTPVPAKKKGVKKRADTTPPTTAAIATSSESSPSTLAAKAK	265
QY	318	LGORESSRPVKKPKKQVDDSOQHAPPEKSSVSQLCCSGILKEMAKHAATWPFY	377
DB	266	IPARESGRPKPKKDLDDSOQHOT-SKKGKLSQLKYCNGILKELSKHAAVAMPFY	324



[illegible]

RESULT 4  
T22847  
hypochemical protein F57C7.1b - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
C:Accession: T22847  
R:White, S.  
submitted to the EMBL Data Library, February 1996  
.;Reference number: Z19625

A:Accession:TT22847  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1087 <MTL>  
A:Cross-references: EMBL:269646; PIDN:CAA93475.1; GSPDB:GN00028; CESP:F57C7.1b  
A:Experimental source: clone F57C7  
A:Genetics:  
A:Gene: CESP:F57C7.1b  
A:Map position: X  
A:Introns: 262/3; 351/3; 391/3; 627/3; 795/2; 957/1  
C:Superfamily: bromodomain homology  
F:307-364/Domain: bromodomain homology <BR01>  
F:579-636/Domain: bromodomain homology <BR02>

Query Match	Similarity	16-7%	Score	608.5:	DB	2:	Length	1087;
Matches	195;	Conservative	115;	Mismatches	261;	Indels	203;	Gaps
								22;
Qy	19	DGLETSOMSTTQAOAOPAPANAASNP	-----	PETSNP-NKPKROTNL	63			
Db	228	DECNDEVEAASNGTEPERGDKADTVNPMWETPRQEPDVGVOVQPTTEPRAGKPTHTNLT	287					
Qy	64	OYLLRVYLKTKMKHOFAPRQOPDAVKLNPDYKIIKTPMDQGTIKKLENNYVMAQ	123					
Db	288	DEVLFTYVKADLKHNHSPFOLPVDAIKLEIPEYHNIVTPRMDLTIEKRLNLYWCAE	347					
Qy	124	ECIOFNFMFNCYLYUNKRGDITVMAAELEKFLYLOKINLEPTREELINVOAKRG	180					
Db	348	DAIKLNLTLFNDCKKFNDRNDIITMCENIBGVVOGRLEMBPSEKPADLADHHRGMS	407					
Qy	191	-----	RGKRETAGKPGVTVP	-----	NTTOASTPQOTOPRN	214		
Db	408	ASTPGKTKGPKQGRKSTGRKRTSLRP-GASTFKDESVDNDNDESSKVDNINFE	466					
Qy	215	PRPVQATGHP-PRATPRLIYQTPVTVVPRQLTPRPVPRQOPRPARAQVQSHNP	273					
Db	467	DTATEKDAESRAESVQPEVNE	-----	EPKPSQOPR	498			
Qy	274	IIAATPOVKTGKGVKRRADT	-----	TPPTIDPIHEPSLPRPEKTKLGLORRES	324			
Db	499	-----	SKKRAAENGERPAVAKINPAI	-----	PVAHPQDET	530		
Qy	325	SRPAKPRKKNDVDS	-----	QOHRAPEKSSKVSSEOLKCCSGILKEMFAKNAAYAMPYKQVD	381			
Db	531	-----	PRKPNPNLTIEKKHLVPRMOGKIRPEWOKFCSKILNIEHVKNGKGAQVFTLLP	584				
Qy	382	VEALGHLHYCDILKHPMSTFKSLREXYDAQDFGADVLMNSCYKVPYRPHVEYA	441					
Db	585	PIKLTITUYLETITNPMQDQTKKRLDKQYAEPEFVNDILAMDNCKYKPKSPAHIS	644					
Qy	442	MARKLQDFEMKFAKMPDEPEEYVA	-----	VSSPAVPRPTKYVAPSSSDSSDSSDSS	498			
Db	645	NALIELRSFEQGRWKLPFRPGVDPIIADSTYINOLN	-----	VNDDLIEDKELNSTL	694			
Qy	499	STDSEERRAORLAELOLKAHVHOLAALSOPQ	-----	ONPKKREKDKK	544			
Db	695	SAVAAEKKCC	-----	AEKLEQTRMSGLTYTIAMORREAKLAGNTAPALANOLSOLEGLGI	751			
Qy	545	EKKREKHKRKEVEE	-----	KKTKKNNSSSNVS	584			
Db	752	SIKSTPMATPELISALSVSSSRAPVPKIIDIIDPSPITAKKIRISKPRASMASNSVPTT	811					
Qy	585	-----	KKEPAPMKSKPPPTYESSEEDCKP	-----	MSYEERQSL	620		
Db	812	PTGNPRGRKPKSGRPKNIVSPATVBSRIQOPFEVTPPVYSIGRVKVVGNDIETLSE	871					
Qy	621	DINKLPGKLGKRVHIIT	-----	SRPSLKNSPDEIIDEFLAKSTIREL	667			
Db	872	RMASIPOMYITPILKLTQIGHNNTGKTTLSKST	-----	CEEEIIDEFNVDSDLVLET	923			

RESULT 5  
T22845





C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002  
 C/Accession: S67605  
 R/Bloescher, H.; Brandt, P.  
 submitted to the Protein Sequence Database, July 1996  
 A/Reference number: S67587  
 A/Accession: S67605  
 A/Molecule type: DNA  
 A/Residues: 1-638 <BLO>  
 A/Cross-references: EMBL:674119; NID:q1431080; PIDD:GMA98636.1; PID:e253212; PID:q143108  
 A/Experimental source: strain S288C  
 C/Genetics:  
 A:Gene: SGD: BDF2; MIPS:YDL070w  
 A/Cross-references: SGD:S0002228  
 A/Map position: 4L  
 C/Superfamily: unassigned bromodomain proteins; bromodomain homology  
 F/158-215/Domain: bromodomain homology <BRO1>  
 F/345-402/Domain: bromodomain homology <BRO2>

Query Match 13.6%; Score 495.5; DB 2; Length 638;  
 Best Local Similarity 22.6%; Pred. No. 2.5e-15;  
 Matches 161; Conservative 117; Mismatches 197; Indels 237; Gaps 18;

```

QY 1 MSASGPGTR-----LNLPYMGDLSTQMTQAQAPQAPANAASSTNPPTSMP 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 VSIDDPHRRDIFHYGHEEYKLAASGI-TNLSSSHAGTSLPISSTNSTEPSEPHR 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 NKPKROT-----NQLYLRLVLTLMKHOFPQOPVDAVKLNLPDY 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 LGIERETPALAEAMEAEELPPHQSKYLLSSI-KATKRLKADAPRLKVPDIALNIPHYF 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 KIITPMQGTIKRLNENYNAQECIDFNFMFNCTYNNKPGDDIVYMAALEKFL 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 NYVOTPMDSLETQKLGQVNVHSVEQVTSDFKTMVNCNLFNFPSSISSMAKIKYFE 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 OKINELPTEETELMIYQAKGRGRKRETKAGVSTVPTQASTPQOTQOPMPVY 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 KTKSNAP----- 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 QATPPPPAVTPDLIVQTPVMTVPPPOPLOTPPPPPQPPAPAPQVOSHPPITIAAT 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 -----PVLPAALK----- 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 PQVTKKGGKRAKADTTT-----TTIDPIHEPPLPPEPKTKLQGRRESS----- 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 -----KTSNRKKKEDMDSPVLRISVSTINDNGES-----GNREGVSGGRPK 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 326 RPYKPPK-KDVPDSQOHPAPEKSSKYSEQLKCCSGILKEKFAKHAAYAMPFYKPYDVEA 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 RTIHPPKSDLPDIYENSKP-KSKTLQKKFRCTKLTKVLMSSKKNNDINPFIQPDPIA 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 385 LGLHLYCDIILKHPMDSTIKSLKAREYRDAQFGADVRLMSNCKYKYNPDHEVYAMAR 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 LNPENFEDYVKKNPMDLGTTISNNLMNKKYKIDPFVDLNVFNCQFNPEGNEVSMGK 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 445 KLQDVEFMFAKMPDEEPYVAVSSPAVPPTKVVAPSSSSSSSSSSSSSSSSSSDDSE 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 KKELELNEHMLEMODILNE-----IEFDSLEEDNYSSTYSDDSD 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 505 EERAQR-----LAELOLKAHVHOLAALSOPOQNKPKKKKKKKKKKKKKKKKEEV 557
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 DEEDINNDITNPAIOYLEQKTKMEVEL-----QQLKQELSKLSKERR-KHLGKTL 508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 558 EENKSKAKPEPPKTKKNNSSNSVSKKEPADMKSPPTYESSEERDKCPMSYEEKRO 617
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 R-----RKAMKHSVD-----DLKKS 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 618 LSLDINKLGEKIGRYVHIIOSSRPS--LKNSPDEIEIDFETLKSTYREL 667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 IYDKINELSDLEKNGMIRITKNSLPADIEITSNDEIEIDLIDLEATTARI 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9  
 T42517

bromodomain protein - fission yeast (Schizosaccharomyces pombe) (fragment)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 02-Sep-2000  
 C/Accession: T42517  
 R/Yoshitoka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
 DNA Res. 4, 363-369, 1997  
 A/Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.  
 A/Reference number: Z17323; MUID:98162722; PMID:9501991  
 A/Accession: T42517  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-361 <YOS>  
 A/Cross-references: EMBL:D89157; NID:q1749521; PIDD:BA13819.1; PID:q1749522  
 A/Experimental source: strain PR745  
 C/Superfamily: bromodomain homology  
 F/64-121/Domain: bromodomain homology <BRO>

Query Match 10.1%; Score 368.5; DB 2; Length 361;  
 Best Local Similarity 28.3%; Pred. No. 6.4e-10;  
 Matches 97; Conservative 54; Mismatches 99; Indels 93; Gaps 8;

```

QY 322 RESSRPVKKPKKDVDPDSQOHPAPEKSSKYSEQLKCCSGILKEKFAKHAAYAMPFYKPYD 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 RETHMDLKHRR-----KDAEKKFQGSYLKELKKQHEKAYAVPFYKPYN 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 VEALGLHYCDIILKHPMDSTIKSLKAREYRDAQFGADVRLMSNCKYKYNPDHEVYA 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 PFAAGCPDYFKYVHKHMDIGTQMKNLNHNEYASMKAFEDAVMLFKNCYKFNAGAPVHL 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 442 MARKLDVFMFAKMPDEEPYVAVSSPAVPPTKVVAPSSSSSSSSSSSSSSSSSSDSS 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 MGRKLESTFOKLMANKPDDSETYMGMS-----VNDYYGYGN 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 499 STDSEERAQRLAELOLKAHVHOLAALSOPOQNKPKKKKKKKKKKKKKKKKKKEVE 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 EYFDSGDE--FLEDGGEFEAVNRQI-----HKLOSTIQ 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 ENKSKAKPEPPKTKKNNSSNSVSKKEPADMKSPPTYESSEERDKCPMSYEEKRO 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 -----AMKSRASSSVSRRSRSLSDVDPPIYTYEQNE-----L 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 SLIDINKLGEKIGRYVHIIOSSRPSLKNSPDEIEIDFETLK 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 AEQCNLTSLADQLSHVAELILKALPHLRNT--DEIEIDVSAMP 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10  
 D96757  
 hypothetical protein T18K17.19 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: D96757  
 R/Theologos, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.; Alon  
 Cln, C. W.; Chung, M. K.; Conn, L.; Conway, A. B.; Conway, A. R.; Creasy, T. H.; Dewar,  
 anen, N. F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J. L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C. A.; Li, J. H.; Lin, X.; Lin, S. X.; Liu, Z. A.; Luros, J. S.; Maiti, R.; Marzia  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Frazer, C. M.; Venter, J. C.; Davis, R. W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: AB6141; MUID:21016719; PMID:11130712  
 A/Accession: D96757  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-461 <STO>  
 A/Cross-references: GB:AE005173; NID:96598866; PIDD:AAF18720.1; GSPDB:GN00141  
 C/Genetics:  
 A:Gene: T18K17.19  
 A/Map position: 1

Query Match 9.3%; Score 339.5; DB 2; Length 461;



Best Local Similarity 26.7%; Pred. No. 1.6e-08;  
Matches 103; Conservative 44; Mismatches 102; Indels 137; Gaps 10;

```
OY 342 PAPERSSKVS-----EOLKCCSGILKEKFAKHAAYAPVPE 383
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 90 PVPKAKLTANGKGGKGGVHGAADKGYOILKSCNNLTFLM--KHKS-GMIFNTPVAV 146
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 384 ALGLHDYCDIIRHMDSTIKSLKAREYDAOFADVRLMFSNCRYPNPDHEVAMA 443
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 147 TLGLHDYHNIIRKPEMDLCTVTRLSKSLKSPLEADVRLFNNAMLNPGVGHVYMA 206
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 444 RKLDVPEMRKAMPDEE-----EPVAVSSPA-----VPPPTKVAVPS 484
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 207 EILNLFEEKWVPLETOYELLIRKOQVPRDIDFNAPVSTNHNVEALRLPAPTSLSP- 265
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 485 SSDSSSDSDSDSDSDSEERAQRLAEOLKAVHQAALASQPQNNPKKEXDK 544
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 266 ----- 265
OY 545 EKKKEKHKEEVEENKSKAKPEPPK---KTKNNSSNSNVSKEPAPKSKPPPYE 600
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 266 -----PPKVENRTLERASMTNPNKPAVLAV-----VPEKL 298
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 601 SEEDKCKPMSEYEERKQSLDINKLPGEKLGRRVHHIISREPSLKNSPDEIDEFELK 660
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 299 VEEASANDLTFDEKROLSEDLQDLPYDKLAVALIIRKRPESLQOD-DEIETDIDSLD 357
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 661 PSTLRELGALCHLFAEKEETFKLR 686
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 358 LETIMEL---FRVTEYKESLSKK 379
    | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 11

T49984  
bromodomain protein-like - Arabidopsis thaliana  
N:Alternate names: protein F12B17.100  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Sep-2000  
C:Accession: T49984  
R:Bevan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25026  
A:Accession: T49984  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <REV>  
A:Cross-references: EMBL:AL353995; GSPDB:GN00063; ATSP:F12B17.100  
C:Genetics:  
A:Gene: ATSP:F12B17.100  
A:Map position: 5  
A:Introns: 73/3; 560/3; 629/1; 656/2  
C:Superfamily: bromodomain homology  
F:269-326/Domain: bromodomain homology <BRO>

Query Match 9.3%; Score 339; DB 2; Length 678;  
Best Local Similarity 25.5%; Pred. No. 2.4e-08;

Matches 105; Conservative 45; Mismatches 112; Indels 150; Gaps 10;

```
OY 314 KTKLQGRRESSRVPKPKKVPDSQOHAPAEKSK--VSQOLKCCSILKEMFAKKHAA 371
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 211 KKKKIGGQKSNPF---ATDEPSLKRHVALDLSEKVLKSMATTCGQILVKKLKH-- 264
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 372 YAMPYRVPVDEALGLHDYCDIIRHMDSTIKSLKAREYDAOFADVRLMFSNCRK 431
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 265 MSWFLNRPVVDVGLHDYHNIIRKPEMDLCTVTRLSKSLKSPLEADVRLFNNA 324
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 432 YNPPDHEVAVAAKRLQDVENR----- 454
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 325 YNPGQGVYLAIAEKILSQFVWFNPTLKRFAQEVKVMGSSSRPGEDNQKVMNNVAE 384
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 455 -----AK-----MPDEPEPVAVAVSSPAVPPPTKVAVAPSSSDSSSDSSD 495
    | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 365 NARKGPEOISIAKKLDYSKPLLPPLPPVIEITRDPSPSPSPVQPPPPSP- 438
OY 496 SSDSTDDSEERAQRLAEOLKAVHQAALASQPQNNPKKEXDKKKEKKEKKEKRE 555
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 439 -----QPVNOVEASL-----E 449
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 556 EYEENKSKAKPEPPKTKRNNSSNSNVSKEPAPKSKPPPYESEEDKCKPMSEYE 615
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 450 VRETNGKRGKGLPKPKADPN-----KRE-----MTMDEK 479
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 616 RQSLDINKLGEKLGRRVHHIISREPSLKNSPDEIDEFELKPTLREL 667
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 480 KGLVNLQELPEPKLGQILQILRKTRDLPQDG-DEIETDIDALNETIMEL 530
    | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 12

A86198  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A86198  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltli, R.; Marzla Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86198  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-766 <STO>  
A:Cross-references: GB:AE005172; NID:98844128; PIDN:AF80220.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 8.9%; Score 323.5; DB 2; Length 766;

Best Local Similarity 24.1%; Pred. No. 1.3e-07;  
Matches 104; Conservative 55; Mismatches 130; Indels 143; Gaps 10;

```
OY 272 PPIATPQPV-----KTKGKVRKADTTPTTIPRHPESLPPREKTKLQGRRE 323
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 339 PREVIARAPVNOISLYLENTQGVNEHVE-----KEKPKKANOYR 381
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 324 SSRPYKPPKDVPSQOHAPAEKSKV-----SEOLKCCSILKEMFAKKHAA 371
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 NSEFLGDK--LPPAESNKKSKSSKKGGVHGFGAGTVKFKCSALLERLKH-- 437
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 372 YAMPYRVPVDEALGLHDYCDIIRHMDSTIKSLKAREYDAOFADVRLMFSNCRK 431
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 438 -GWVFNAVPDVGLGLDYTYTIEHPMDLGTIKSALMKNLKSPREFAEDVRLTFHNA 496
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 432 YNPPDHEVAVAAKRLQDVEMFAKMPDE-----PEEPVAVASPAV-PP 475
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 497 YNPGQGVYLAIAVTLQIFERMAVIEADYNREMFVGYEENLPTPKRSKLGTMPP 556
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 536 PTKVAVPSSSDSSSDSDSDSDSEERAQRLAEOLKAVHQAALASQPQNNPK 535
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 557 PINVFNITDRADWSNRQPTTTPGRTSATPSG-----RTPALKKRANE 601
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 536 PKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 595
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 602 PNKR----- 606
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 596 PPTYSEEDKCKPMSEYEERKQSLDINKLPGEKLGRRVHHIISREPSLKNSPDEIDE 655
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 607 -----MTYEKQKLSGHLQNLPPDKLDAIVQVNNKNTAVK-LRDEIEVD 651
    | | | | | | | | | | | | | | | | | | | | | | | |
OY 656 FETLKPSLREL 667
    | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 652 IDSDVDETLWEL 663

# RESULT 13

S39161

CREB-binding protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 02-Aug-2002

C:Accession: S39161

R:Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Haglwaara, M.; Montminy, M.R.; Goodman, R.H.

Nature 365, 855-859, 1993

A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.

A:Reference number: S39161; PMID:94019866; PMID:8413673

A:Accession: S39161

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2441 <CHR>

A:Cross-references: GB:S66385; NID:9435854; PIDN:AAB28651.1; PID:9435855

A:Superfamily: Transcription coactivator CREB-binding protein; bromodomain homology

F:1112-1169/Domain: bromodomain homology <BRO>

Query Match 8.9%; Score 323.5; DB 2; Length 2441;

Best Local Similarity 23.4%; Pred. No. 4.2e-07;

Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

Db 9 TRLRLPVNGDGLTSMSTTQAQAPQANAASTNPP--PETSNDPKKQTNOLQYL 66

670 TRLHKGILGN-----QPALPASGAQPVIPPAQSVPRPN----- 704

Qy 67 LRVYLKTLKHPMPQPVDAVKLN--LPDYKILITPKMDGTLK-----KR 113

Db 705 -----GLPLRPVNMQVSGQNSR-----NPMSLGNVQLPQAPMGPRASP 745

Qy 114 LENNYWNAOECI-----ODENFMETNCYIYKRGDDIVLAEAL-LEKLFLOKI 161

Db 746 MNHSVQMSASYPGMAISPSRMPORPNMGTIA--NN-----TMAQAPTONQIFPQ- 795

Qy 162 NELPTEETELMIVQAKRGGRKGTGA--KPGVS-----TVPNTTQASTPPROTTP 211

Db 796 NOPPS--SSGAMSVNAGMCPAAGVSGQEPGALPNLMLAPQASQLPCPPVYQSP 854

Qy 212 -QPNPP-----VQATPRP--FPAVTPDLIVQTPVMTVPRQL--QTPPPPP- 254

Db 855 LHPPTPASTAGMPSLDHHTAPGMP--PQPAAPTOSTVSSGQTPPTPPGSVPSA 910

Qy 255 -----PQPPPPAP-----APQVQSHPRPI- 275

Db 911 AQTOSTPTVQAAQAQVTPQPTQPTVQPPSVATPQSSQOQPTVHTQPTGRLSQAAASID 970

Qy 276 --AATPQPVKTKGKVRKADTTTP-----TTIDPHEPRLPPRPKTKTLGQRRESS 325

Db 971 NRPPTPTVTSAPTSSQGPDPVMLEMKTEVQTDAPETPESKRESEMEEDLOGS 1030

Qy 326 RPKV-----PKKDVDSQHPAPE-----KSSVSVSOLKCCSILTEMKRAK- 368

Db 1031 SOVKETDTTEQSEPEVEVEKKPEVKEAKEEENSSNNTAQSSTSPSPRRKIKRPEE 1090

Qy 369 -----HAAY-----AMPYKPVDEALGLHDYCDIIKHPMDSTIKSLAEAYRD 414

Db 1091 LKQALMPTLEALYROPDESIFRQVPDPQLGIDYFDIYKPNMDLSTIKRLDTGQYOE 1150

Qy 415 AOEFGADVRLMFSNCKYKPPDPHEVVAAMARKIDVFE 451

Db 1151 PMQYVDVDRIMFNAMLYNKRKTSRYVKFKCSKLALEVE 1187

# RESULT 14

H86312

P2H15.2 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001

C:Accession: H86312

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

anssen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marcia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: H86312

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <STO>

A:Cross-references: GB:AEO05172; NID:99665057; PIDN:AAF97259.1; GSPDB:GN00141

A:Genetics:

A:Map position: 1

Query Match 8.7%; Score 318; DB 2; Length 440;

Best Local Similarity 27.0%; Pred. No. 1.4e-07;

Matches 93; Conservative 46; Mismatches 83; Indels 122; Gaps 9;

Qy 355 KCCSGIIEKFAKKHAYMPFYKPPVDEALGLHDYCDIIRKPMDSSTIKSLAEAYRD 414

Db 134 KNCNSLTTLKLM--KHS-AWFFVVPDAKGLGHDYHNHYKEPMDLGTVTKLGKSLYKS 190

Qy 415 AOEFGADVRLMFSNCKYKPPDPHEVVAAMARKIDVFEEMKRAKPDPE----- 462

Db 191 PLDFAEVDRITFNAILIYNFIGHDYTRFALLNMEEDKKNVSTEMOYDNIHKKFKPTRI 250

Qy 463 -----EPVAV-----SSPAPPTKYVAPPSSSSSSSSSSSSSSSSSSSS 503

Db 251 EPPAPAPSIAPVEIPALVPSPPSPPPPPPPVPAAPVLEN----- 293

Qy 504 EERRAQRALQOLAKVAHQLAALSOPOQNKKKKKKKKKKKKKKKKKKKKKKK 563

Db 294 -----RTWERESMTIPVEPAVITAP----- 316

Qy 564 KAKEPPPKTKTKNNSSNSVSKKEPAPMKSPPTVESEEDCKPMSEKRLSDIN 623

Db 317 KAFE-----EAPVNN-----RLTLEKRLSEELQ 343

Qy 624 KLPEKLRGVNHIIOSEKPSLKNNDPEIETETKPTLREL 667

Db 344 LPEYKLETVVQIIRKSNPESQKD-DETELIDSDIDINTIMEL 386

# RESULT 15

S39162

transcription coactivator CREB-binding protein - human

N:Alternate names: CBP; RSTs; Rubinstein-Taybi syndrome (RSTS) protein

C:Species: Homo sapiens (man)

C:Date: 07-Oct-1994 #sequence\_revision 17-Nov-1995 #text\_change 02-Aug-2002

C:Accession: S39162; S60345; I58096

R:Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Haglwaara, M.; Montminy, M.R.; Goodman, R.H.

Nature 365, 855-859, 1993

A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.

A:Reference number: S39161; PMID:94019866; PMID:8413673

A:Accession: S39162

A:Molecule type: mRNA

A:Residues: 1-2440 <CHR>

A:Note: differences with the mouse sequence are shown

R:Lundblad, J.R.; Kwok, R.P.S.; Laurence, M.E.; Harter, M.L.; Goodman, R.H.

Nature 374, 85-88, 1995

A:Title: Adenoviral ELA-associated protein p300 as a functional homologue of the tran

A:Reference number: S60344; PMID:95174889; PMID:7870179

A:Accession: S60345

A:Status: preliminary

A:Molecule type: protein

A:Residues: 'S', 574-670, 'T', 672-681 <LUN>

R:Pettili, F.; Giles, R.H.; Dawerise, H.G.; Saris, J.J.; Hennekam, R.C.; Masuno, M.; T

Nature 376, 348-351, 1995

A:Title: Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator  
A:Reference number: 158096; MUID:95336817; PMID:7630403  
A:Accession: 158096  
A>Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 352-356 <PEP>  
A:Cross-references: GB:089354; NID:91888537; GB:S78936; NID:91041931  
A:Note: this translation is not annotated in GenBank entry S78936, release 112.0  
A:Note: this sequence with a termination mutation is from a patient with Rubinstein-Taybi  
C:Genetics:  
A:Gene: GDB:CREBBP; RTS; CBP; RSTS  
A:Cross-references: GDB:437159; OMIM:180849; OMIM:600140  
A:Map position: 16p13.3-16p13.3  
A:Note: defects in this gene may result in Rubinstein-Taybi syndrome  
C:Superfamily: transcription coactivator CREB-binding protein; bromodomain homology  
C:Keywords: phosphoprotein; transcription; zinc finger  
F:462-661/Domain: CREB binding #status predicted <CBP>  
F:111-1168/Domain: bromodomain homology <BRO>  
F:1283-1311/Region: zinc finger CCCC motif  
F:1707-1733/Region: zinc finger CCCC motif  
F:78,381,745,1172,1533,2063,2354/binding site: phosphate (Ser) (covalent) (by calmodulin  
F:1771/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted

Query Match 8.58; Score 310; DB 2; Length 2440;  
Best Local Similarity 25.28; Pred. No. 1.7e-06;  
Matches 105; Conservative 54; Mismatches 133; Indels 124; Gaps 17;

QY 148 LMAEA-LEKLELQKINELPTEETIMIVQAKRGKREKTAK--PGVSTV----- 196  
DB 783 MMAQAPASQOFLPQ-NQFPSSGAMSV---GKQPPAQGTSGQVPGALPBNPLNMG 837  
QY 197 PNTTQASTPQGTQTP-QPNPPVQATPHPPAVTPDLIVQTVMTVVPPPL----- 247  
DB 838 PQASQLPCPPVYQSLPHTPPASTA-----AGMPSLQHTTP-PGKMPDPQAPATQPSSTP 891  
QY 248 -----QTPPPVP-----PQP-----QPPAPAPQVQSHP-P 273  
DB 892 VSSSGQTPPTPGSVPSATQSTPTVQAAQAQVTPQPPQPPSVATPQSSSQOQPTP 951  
QY 274 IIAATPQPVKTKGKVRKRAADTTPTTI-----DPIHEPP 307  
DB 952 VHAQPTPTPLSQAAASIDNRPTSTYASATNSQOQPPDVPLVLEMKETQAEDETPDPG 1011  
QY 308 SLPEPPTTKLGQRRESSRPVKKKDVPSQHP-----APEKSSKVSQLKCCSG-- 359  
DB 1012 ESKGEPRSEMMEDLQASQVK-ETDIAEQSRPMEVEDKKPVKVEVKEEESSSNGT 1070  
QY 360 -----LIKEMFAKKHAY-----AMPFYKPYDVEALGLHIDYCDIY 395  
DB 1071 ASQSTSPSQPRKIKFKEELRQALMPTLEALYRQDPESLPRQPPVDPLGIPDYFDIVK 1130  
QY 396 HPMDSTIKSLAEAREYDAQAEFGADVILMFSNCYKYNPPDHEVYAMARKLDVFE 451  
DB 1131 NPMDLSTIKRKLDRGQVQEPQVYDDVWLMFNMMALYNRKTSRYKKCSKLAIEVFE 1186

Search completed: February 25, 2003, 13:22:44  
Job time : 55 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:13:46 ; Search time 14 Seconds  
(without alignments)  
2038.265 Million cell updates/sec

Title: us-09-700-590A-22

Perfect score: 3639

Sequence: 1 MSAESGPTRLNPLVMGDS...ALCHLLFAEEKETFKRLKLM 688

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3531.5	97.0	1362	1 BRD4_HUMAN	O60885 homo sapien
2	1910.5	52.5	726	1 BRD3_HUMAN	Q15059 homo sapien
3	1888	51.9	801	1 BRD2_HUMAN	P25440 homo sapien
4	1339.5	36.8	2038	1 FSHL_DROME	P13709 drosophila
5	570.5	15.7	686	1 BDF1_YEAST	P35817 saccharomyc
6	323.5	8.9	2441	1 CBP_MOUSE	P45481 mus musculus
7	308	8.5	2442	1 CBP_HUMAN	Q92793 homo sapien
8	305	8.4	2414	1 P300_HUMAN	Q09472 homo sapien
9	300.5	8.3	1484	1 CES2_HUMAN	Q9Bxf3 homo sapien
10	267.5	7.4	1020	1 NFH_HUMAN	P12036 homo sapien
11	265	7.3	2056	1 CBP1_CAEEL	P34545 caenorhabdi
12	262.5	7.2	831	1 NFH_RAT	P16884 rattus norv
13	251.5	6.9	1087	1 NFH_MOUSE	P19246 mus musculus
14	231	6.3	699	1 NP14_HUMAN	Q14978 homo sapien
15	221.5	6.1	817	1 VRP1_YEAST	P37370 saccharomyc
16	218	6.0	1509	1 GSR1_HUMAN	Q9nmz4 homo sapien
17	216	5.9	279	1 Y91_NEPVO	Q10341 orgyia pseu
18	216	5.9	1214	1 BR3_HUMAN	Q04341 homo sapien
19	213.5	5.9	2715	1 TRX2_HUMAN	Q9umh6 homo sapien
20	213	5.9	283	1 EXTN_SORBI	P24152 sorghum bic
21	213	5.9	2295	1 WDR9_HUMAN	Q9nsi6 homo sapien
22	212.5	5.8	555	1 GPI_CHLRE	Q9fpq6 chlamydomon
23	212.5	5.8	1217	1 AF4_MOUSE	O88573 mus musculus
24	211.5	5.8	534	1 APG_ARATH	P40602 arabidopsis
25	211	5.8	1575	1 SYJ1_HUMAN	O43426 homo sapien
26	209.5	5.8	2068	1 T2D1_DROME	P51123 drosophila
27	209	5.7	1050	1 TE1A_HUMAN	O15164 homo sapien
28	208.5	5.7	267	1 EXTN_MAIZE	P14918 zea mays (m
29	206.5	5.7	3421	1 TEGU_HSVB	P28955 equine herp
30	205.5	5.6	819	1 FYB_MOUSE	O35601 mus musculus
31	205	5.6	707	1 SFPQ_HUMAN	P23246 homo sapien
32	204.5	5.6	434	1 N075_LUPLU	O06841 lupinus lut
33	204.5	5.6	1386	1 ZAP3_MOUSE	Q9R017 mus musculus

#### RESULT 1

ID	BRD4_HUMAN	STANDARD;	PRT;	1362 AA.
AC	O60885; Q96PD3;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Bromodomain-containing protein 4 (HUNK1 protein).			
GN	BRD4 OR HUNK1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	French C.A., Fletcher J.A.;			
RT	"Human BRD4 protein."			
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1-722 FROM N.A.			
RC	TISSUE=Placenta;			
RA	Weber B.;			
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	-1- SIMILARITY: CONTAINS 2 BROMODOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; AF386649; AAL26987.1; -			
DR	EMBL; Y12059; CAA72780.1; -			
DR	HSSP; Q92831; 1B91.			
DR	Gene; HGNC:13575; BRD4.			
DR	InterPro; IPR001487; Bromodomain.			
DR	Pfam; PF00439; bromodomain; 2.			
DR	PRINTS; PR00503; BROMODOMAIN.			
DR	SMART; SM00297; BROMO; 2.			
DR	PROSITE; PS00633; BROMODOMAIN_1; 1.			
DR	PROSITE; PS50014; BROMODOMAIN_2; 2.			
KW	Bromodomain; Repeat; Nuclear protein.			
FT	DOMAIN 75 147 BROMODOMAIN 1.			
FT	DOMAIN 368 440 BROMODOMAIN 2.			
FT	DOMAIN 535 594 LYS-RICH.			
FT	DOMAIN 692 717 SER-RICH.			
FT	DOMAIN 703 714 POLY-SER.			
FT	DOMAIN 738 743 POLY-HIS.			
FT	DOMAIN 757 761 POLY-PRO.			
FT	DOMAIN 764 770 POLY-PRO.			
FT	DOMAIN 771 775 POLY-GLN.			
FT	DOMAIN 776 783 POLY-PRO.			
FT	DOMAIN 954 964 POLY-PRO.			

```
FT DOMAIN 974 986 POLY-PRO.
FT DOMAIN 1011 1014 POLY-PRO.
FT DOMAIN 1028 1033 POLY-PRO.
FT DOMAIN 1283 1300 POLY-GLN.
FT DOMAIN 1301 1308 POLY-ALA.
FT DOMAIN 1335 1338 POLY-ARG.
FT CONFLICT 720 721 EM -> GP (IN REF. 2).
SQ SEQUENCE 1362 AA; 152219 MW; D52FEFCF9960907 CRC64;

Query Match 97.0%; Score 3531.5; DB 1; Length 1362;
Best Local Similarity 98.0%; Pred. No. 6.9e-142;
Matches 672; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

Qy 1 MSAESGPGTRLNRLNLPVMDGLTSMSTTQAQAPQAPANAASSTNPPPTSPNPKPKRQT 60
Db 1 MSAESGPGTRLNRLNLPVMDGLTSMSTTQAQAPQAPANAASSTNPPPTSPNPKPKRQT 60
Qy 61 NQQLYLLRVLLTKLWKHFQFAPFPQPDVAVKLNLDYKIKTPMDMTIKRLNNYVW 120
Db 61 NQQLYLLRVLLTKLWKHFQFAPFPQPDVAVKLNLDYKIKTPMDMTIKRLNNYVW 120
Qy 121 NQAECIQDFNTMFTNCYIYNKPGDDIVLMAEALFKLQKINELPTEETIMVQAKRG 180
Db 121 NQAECIQDFNTMFTNCYIYNKPGDDIVLMAEALFKLQKINELPTEETIMVQAKRG 180
Qy 181 RGRKETGAKGVSVVNTQASTPQOTQTPNPPVQATPHPPVATPDLLVQTPVMT 240
Db 181 RGRKETGAKGVSVVNTQASTPQOTQTPNPPVQATPHPPVATPDLLVQTPVMT 240
Qy 241 VVPQPLQTPPPVPPQPPAPAPQVSHPPPIIAATPQPKTKKGRKADTTPTTI 300
Db 241 VVPQPLQTPPPVPPQPPAPAPQVSHPPPIIAATPQPKTKKGRKADTTPTTI 300
Qy 301 DPIHEPPSLPPEPKTKLQGRRESSRPVKKKDVDPDQHPAPKSKVSEQLKCCSGI 360
Db 301 DPIHEPPSLPPEPKTKLQGRRESSRPVKKKDVDPDQHPAPKSKVSEQLKCCSGI 360
Qy 361 LKEMPAKHAAYAMPFYKPDVVEALGLHDYCDIHKHPMDSTIKSLKLEAREYRDAEFGA 420
Db 361 LKEMPAKHAAYAMPFYKPDVVEALGLHDYCDIHKHPMDSTIKSLKLEAREYRDAEFGA 420
Qy 421 DVLRLMFSNCYKYNPDHVEVAMARKLDQVEMRFAMKDEPEEPVAVSSPAVPPTKV 480
Db 421 DVLRLMFSNCYKYNPDHVEVAMARKLDQVEMRFAMKDEPEEPVAVSSPAVPPTKV 480
Qy 481 APPSSSSSSSSSDSDSDSDSEERAEQLAEQELKAVHEQLAALSQPOQNKPKKE 540
Db 481 APPSSSSSSSSSDSDSDSDSEERAEQLAEQELKAVHEQLAALSQPOQNKPKKE 540
Qy 541 KDKKKEKKEKKEVEENKSKAKEPPPKTKKNNSSNSNVSKKEPAPMKSPPTTYE 600
Db 541 KDKKKEKKEKKEVEENKSKAKEPPPKTKKNNSSNSNVSKKEPAPMKSPPTTYE 600
Qy 601 SEEDKCPMSVEERKQLSLDINKLPGKGLGVVHIQSRPPLSKNSNPDEIDFETLK 660
Db 601 SEEDKCPMSVEERKQLSLDINKLPGKGLGVVHIQSRPPLSKNSNPDEIDFETLK 660
Qy 661 PSTRELKALCHLLFAEEKETFKLRK 686
Db 661 PSTRELKALCHLLFAEEKETFKLRK 677
```

## RESULT 2

```
BRD3_HUMAN
ID BRD3_HUMAN STANDARD; PRT; 726 AA.
AC Q15059; O92645;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 3 (RING3-like protein).
GN BRD3 OR RING3L OR KIAA0043.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT analysis of cDNA clones from human cell line KG-1.";
RT DNA Res. 1:223-229(1994).
RL [2]
RN SEQUENCE OF 363-726 FROM N.A.
RP MEDLINE=98038990; PubMed=9373153;
RX Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
RA "Chromosomal localization, gene structure and transcription pattern of
RT the ORFX gene, a homologue of the MHC-linked RING3 gene.";
RL Gene 200:177-183(1997).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D26362; BAA03393.1; -
CC EMBL; Z81330; CAB03630.1; -
CC HSSP; Q92831; 1B91.
CC Genew: HGNC:1104; BRD3.
CC MIM; 601541; -
CC InterPro; IPR001487; Bromodomain.
CC Pfam; PF00439; bromodomain; 2.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 2.
CC PROSITE; PS00633; BROMODOMAIN_1; 2.
CC PROSITE; PS00634; BROMODOMAIN_2; 2.
CC Bromodomain; Repeat; Nuclear protein.
CC DOMAIN 56 115 BROMODOMAIN 1.
CC DOMAIN 326 398 BROMODOMAIN 2.
CC DOMAIN 487 555 LYS-RICH.
CC DOMAIN 676 725 SER-RICH.
CC CONFLICT 465 466 EL -> DV (IN REF. 2).
CC SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 52.5%; Score 1910.5; DB 1; Length 726;
Best Local Similarity 59.9%; Pred. No. 5.2e-74;
Matches 396; Conservative 66; Mismatches 144; Indels 55; Gaps 15;

Qy 27 STTQAQAPQAPANAASSTNPPPTSPNPKPKRQTQLOLYLLRVLLTKLWKHFQFAPFPQ 86
Db 4 ATTAVAP-GIPATPGVNPVPPPEVSNPKGRKTNQLOLYLLRVLLTKLWKHFQFAPFPQ 62
Qy 87 VDAVKLNLDYKIKTPMDMTIKRLNNYVNAQECIQDFNTMFTNCYIYNKPGDDI 146
Db 63 VDAVKLNLDYKIKTPMDMTIKRLNNYVNAQECIQDFNTMFTNCYIYNKPGDDI 122
Qy 147 VLMAEALFKLQKINELPTEETIMVQAKGRGR-GRKETGAKGVSVVNTQASTPQOTQTPNPP 202
Db 123 VLMAEALFKLQKINELPTEETIMVQAKGRGRGRKETGAKGVSVVNTQASTPQOTQTPNPP 180
Qy 203 STTQOTQTP-QPNPPVQATPHPPVATPDLLVQTPVMTVPPVPPVPPVPPVPPVPPVPP 261
Db 181 -TPFQSVPPVTSQTP-----VIAATPVPTITAN---VTSVPVPPVPPVPPVPPVPP 221
Qy 262 APAQVQSHPPPIIAATPQPKTKKGRKADTTPTT-----TIDPIHEPPSLPPEPKTKK 317
Db 222 -----PATPIVPPVPPTP-PVKKKGVKRRKADTTPTTSAITASRSSESPPL-SDPKQAK 274
```







Db 928 RGSKKKPSQVMN-----FSEEDTAKPMYDEKKRQLSLDINKLPDGLGRVHIIQNR 988

Qy 641 EPSLKNSNPDEIEIDFETLKPSTLREL 667  
||||:|||||||||||||||||

Db 983 EPSLRDSNPDEIEIDFETLKPSTLREL 1009

RESULT 5

BDFL_YEAST	STANDARD;	PRT;	686 AA.
ID	BDFL_YEAST		
AC	P35817; Q06048;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	BDf1 protein.		
GN	BDf1 OR YLR399C OR L0804.18.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
[1]			
RP	SEQUENCE FROM N.A.		
RP	STRAIN-S288C;		
RC	MEDLINE=95116323; PubMed=7816623;		
RX	Lygerou Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,		
RA	Sentenac A., Seraphin B.		
RA	"The yeast Bdf1 gene encodes a transcription factor involved in the		
RT	expression of a broad class of genes including snRNAs.";		
RL	Nucleic Acids Res. 22:5332-5340(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	Roeder G.S., Chua P.R.;		
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.		
[3]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288C / AB972;		
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,		
RA	Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,		
RA	Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,		
RA	Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,		
RA	Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,		
RA	Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,		
RA	Wilson R., Waterston R.;		
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.		
[4]			
RP	SEQUENCE OF 471-686 FROM N.A.		
RC	MEDLINE=93309467; PubMed=8321235;		
RA	Widner W.R., Wickner R.B.;		
RA	"Evidence that the SKI antiviral system of Saccharomyces cerevisiae		
RT	acts by blocking expression of viral mRNA.";		
RL	Mol. Cell. Biol. 13:4331-4341(1993).		
CC	-1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE EXPRESSION OF A		
CC	BROAD CLASS OF GENES INCLUDING SNRNAs. REQUIRED FOR SPOULATION.		
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).		
CC	-1- SIMILARITY: CONTAINS 2 BROMODOMAINS.		
CC	-1- SIMILARITY: CONTAINS 1 ET DOMAIN.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commer-		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
DR	EMBL; Z18944; CAA79377.1; -		
DR	EMBL; U18116; AAA89115.1; -		
DR	EMBL; U19729; AAB82357.1; -		
DR	EMBL; L13469; AAA35048.1; -		
DR	PIR; A8140; A48140.		
DR	PIR; S41801; S41801.		
DR	TRANSFAC; T03204; -		
DR	SGD; S0004391; BDF1.		

[illegible]





colon, spleen, kidney.

-1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a developmental disorder associated with the duplication of a 2 Mb region of 22q11.2. Duplication usually takes in the form of a supernumerary bisatellited isodicentric chromosome, resulting in four copies of the region (represents an inv dup(22)(q11)). CES is characterized clinically by the combination of coloboma of the iris and anal atresia with fistula, downslanting palpebral fissures, preauricular tags and/or pits, frequent occurrence of heart and renal malformations, and normal or near-normal mental development.

-1- SIMILARITY: CONTAINS 1 BROMODOMAIN.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF336133; AAK15343.1; .  
 EMBL; AB051527; BAB21831.1; .  
 EMBL; AF411609; AAL07393.1; .  
 HSSP; Q92831; 1B91.  
 Genew; HGNC:1840; CECR2.  
 InterPro; IPR001487; Bromodomain.  
 Pfam; PF00439; bromodomain; 1.  
 PRINTS; PR00503; BROMODOMAIN.  
 SMART; SM00297; BROMO; 1.  
 PROSITE; PS00633; BROMODOMAIN\_1; 1.  
 PROSITE; PSS0014; BROMODOMAIN\_2; 1.  
 KW Alternative splicing.  
 FT DOMAIN 451 521 BROMODOMAIN.  
 FT DOMAIN 333 337 POLY-GLU.  
 FT DOMAIN 611 614 POLY-SER.  
 FT DOMAIN 1250 1253 POLY-PRO.  
 FT VARSPIC 291 318 MISSING (IN ISOFORM B).  
 FT VARSPIC 519 526 EYKMSDN -> GKGRSLC (IN ISOFORM B).  
 FT VARSPIC 527 1484 MISSING (IN ISOFORM B).  
 FT CONFLICT 370 389 MISSING (IN REF. 2).  
 FT CONFLICT 1029 1029 C -> S (IN REF. 2).  
 FT CONFLICT 1045 1045 R -> W (IN REF. 2).  
 SQ SEQUENCE 1484 AA; 164214 MW; 049AA844E51AF63F CRC64;

Query Match 8.3%; Score 300.5; DB 1; Length 1484;  
 Best Local Similarity 20.9%; Pred. No. 4.6e-06;  
 Matches 140; Conservative 85; Mismatches 265; Indels 179; Gaps 26;

QY 47 PETS--NPKNPKRTQVQLYLRV-----VLTKLKHQPAWPPQPPVDAVKLNLP 95  
 DB 412 PPELHLDNPSMKREKTKDLFELDDFTAMVKVLDVVAHKDSWPFLEPVD--ESYAP 469  
 QY 96 DYKIIITPMDMGTIKKLENNYVNAQECIQDFTMTNCTIYNKPGDDIVLMAEALEK 155  
 DB 470 NYQIIKAPMDISSMEKGLNGLYCTEEFVNDMKTFRNCRKINGESSEYTKMSDLNR 529  
 QY 156 LFQ-KINELPTE---ETEMIVQAKGRGRKKTGA----- 189  
 DB 530 CFHAMKHPFGEDGDTDEEFWIREDEKREKRSRACRSGSHVWTRSDRPEGSSRRQQP 589  
 QY 190 -KGVSTVNTQASTPQTPQPPPPVQ-----ATPHPP-ATVPDLIVQTPVMT 240  
 DB 590 MNGGKSLPPTTRAPSSGDDQSSSTQPPREVGTSGRGFSPLHLCGGTGS---QAPFLN 646  
 QY 241 VVPPQLQTPPPVP---PQQPPPPA-----PAP-QPVQSHPIIAATPQPVYTK 285  
 DB 647 -----QMPAPVGTGRLGSDPAFLYSSGVLEPHGPEVQQRQPPFTMQPPVGNLS 699  
 QY 286 KGVKRAADTTTPTTIDPIHEPPSLPPEPKTKLQ-----RRE 323  
 DB 700 RG-PRLTPEEKQCGGLTHLSNNGPHPGSLQLQISGPDQSMYAPAQFQGFIPRH 758

QY 324 SSRPVKPKVDVPSQQHAPKESKSVSEQLKCCSGILKEMFAKKAAY----- 372  
 DB 759 GGAPARPP--DFESSEIP-----PSHMYRSYKILNRVHSAVNGNHGATNGPLGPDEKP 812  
 QY 373 ---AWPFYKPVVDVREALGHDYCDIHKHPMDMSTKSKLEAREYRDAQEF---GADVRLM 425  
 DB 813 HLGPGPSHQP---RTLQ-----HYMDSRVMRPPPPNQMTQSGFLPHGVPSSGYM 860  
 QY 426 FSNC-----YKNPP-----DHEVVAMARKLDQVFEFR 454  
 DB 861 RPPCKSAGRLQPPVPAPSSLFGAPALRGVGGDSMDSPEMIAMQQLSSRVCPGCV 920  
 QY 455 AKMDEPEEPVAVSSPAVPPTKVAPSSSDSSSDSSSDSDSDSEERARQRLAEL 514  
 DB 921 PYHPQAPRLPLPGFPQVAHPMSTVSAPKALGNPCRPENSEAQEPENDQAEPLPGL 980  
 QY 515 QEOLKAVHEQ---LAALSQP-----QONPKKKEKDKKKKKKKHKKKEVEZNK 561  
 DB 981 EEPGPGVTSGVYLTQLPHPTPLQTDCTROSSPOERTVGPPELKS---SCSESADNCK 1037  
 QY 562 KSKAKEPPPP 570  
 DB 1038 AMGKNPREP 1046

RESULT 10  
 NFH\_HUMAN  
 ID NFH\_HUMAN STANDARD; PRT; 1020 AA.  
 AC P12036;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
 DE (Neurofilament heavy polypeptide) (NF-H).  
 GN NEFH OR NEFH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88328981; PubMed=3138108;  
 RA Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;  
 RT "The structure and organization of the human heavy neurofilament  
 subunit (NF-H) and the gene encoding it.";  
 RL EMBO J. 7:1947-1955(1988).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NEF IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X15306; CAA33366.1; .  
 EMBL; X15307; CAA33366.1; JOINED.  
 EMBL; X15308; CAA33366.1; JOINED.  
 EMBL; X15309; CAA33366.1; JOINED.













```
Qy 54 NKPRQTNQLOYLRLVVLKTLWKHQFAPFPQOPVDAYKL-----NLPDYKIKITPMD 106
Db 134 PIPNAPLS-----PAPVPSIPSSAPPIDIPSSAAPPIP 169
Qy 107 MGTIKKRLNNYYWNAQECIQDFNTMETNICYIYNKPGDDIVLMAEAKLFLQKINELPT 166
Db 170 I-----VPSSAPPPLSLGASAPK-----VPQ 191
Qy 167 EETIMIVQAKGRGRGKGTAKRGVSTVP-----NTQASTPPQTQTPQPNP----- 215
Db 192 NRPHMPSVRPAHRSHQRKSSNISLSVSAAPLPSASLPTHVSNPPQAPPPPTTIGLIDS 251
Qy 216 -----PP-----VOATPHFPFAVTPDL 232
Db 252 KNIKPTDNVSPSPSEVPAGGLPLAEINARRSERGAVGVSTKIOTENHKSPS----- 306
Qy 233 IVQTPVMTVVPPQLOTPPPVPPQPPPP-----APAPQPVQSHPPITIAAT 278
Db 307 --QPPLPSSAPPITSHAPPLPPTAPPPLPPSLPNVTSAPKKATSAPAPP-----PPLPAA- 359
Qy 279 PQPVTKKGVKRKADTTPTTTIDIHPPSL-PPEPKTKLQORRESSRVKPKKDVDP 337
Db 360 -----MSSASTNSVKATEV--PPTLAPPLPNTTSVPPNKAASSMPAPPPPP----- 402
Qy 338 SQHPAPEKSKSVSEQLKCCSGILKEMFAKHAAYANPFYKPDVDEALGLHDYCDIHKP 397
Db 403 -----PPPGAFSTSSALSASSIPLAPL----- 425
Qy 398 MDMSTIKSLEAREYRDAQEFGADVRLMFSNCYKYNPPDHEVWAMARKLODVFEMRFKM 457
Db 426 -----PP----- 427
Qy 458 PDEPEPVAVSSPAVPPPTKVA---PPSSSDSSSDSSDSDSDTDDSEERAQRLAEL 514
Db 428 ---PPPSVATSVPSAPPPTLTNNKPSASSKQISSSSSSSAVTPGGP--LPFLAEI 482
Qy 515 QEOLKAVHEQLAALSQONKPKKKEKKKKHKKKEEVEENKSKAKEPPPKTK 574
Db 483 QKK-----RDRFVVGDTGYTTQDKQEDVIGSSKDDNVRPSPISPSIN-----PPKQSS 532
Qy 575 KNNS-----NSNVSKK-----EPAPMKSKPPPTVESEED 605
Db 533 QNGMSFDETESKLHKQTSNAFNAPPHTDAMAPPLPPSAPPPTTSLPTPTASGDDHT 592
Qy 606 KCKPMSTEERQLSLDINK-----LPG 627
Db 593 -----NDKSETVLGMKKAKAPALPG 612
```

Search completed: February 25, 2003, 13:21:03  
Job time : 23 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:17:31 ; Search time 41 Seconds  
(without alignments)  
3457.571 Million cell updates/sec

Title: US-09-700-590A-22  
Perfect score: 3639  
Sequence: 1 MSASGPGTRLNLPVMDG.....ALCHLLFAEKEKTFKRLKLM 688

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3531.5	97.0	731	4	O60433
2	3397	93.3	723	11	Q8VHF7
3	3392	93.2	1400	11	Q9ESD6
4	3392	93.2	1400	11	Q8VHF8
5	1915	52.6	726	11	Q9J125
6	1901	52.2	798	11	Q88411
7	1901	52.2	798	11	O54795
8	1887	51.9	801	4	Q969U4
9	1876.5	51.6	729	13	Q90971
10	1757	48.3	701	13	Q8QF77
11	1690.5	46.5	814	13	Q8UUM2
12	1680	46.2	664	5	Q8T775
13	1678	46.1	732	5	Q9GU61
14	1540	42.3	956	11	Q91Y44
15	1501.5	41.3	947	4	O14789
16	1401	38.5	503	11	Q99PC5

17	1339.5	36.8	1937	5	Q9W3L3	Q9W3L3 drosophila
18	763.5	21.0	851	5	Q95F80	Q95F80 caenorhabdi
19	608.5	16.7	1087	5	Q20948	Q20948 caenorhabdi
20	600.5	16.5	1250	5	Q20947	Q20947 caenorhabdi
21	561	15.4	578	3	Q9Y7N0	Q9Y7N0 schizosacch
22	531.5	14.6	727	3	Q9HGP4	Q9HGP4 schizosacch
23	498.5	13.7	249	13	Q91686	Q91686 xenopus lae
24	495.5	13.6	638	3	Q07442	Q07442 saccharomyc
25	430	11.8	513	5	Q9VCG6	Q9VCG6 drosophila
26	375	10.3	507	10	Q9LHY2	Q9LHY2 oryza sativ
27	368.5	10.1	361	3	P78808	P78808 schizosacch
28	365.5	10.0	128	11	O35692	O35692 mus musculu
29	351.5	9.7	370	5	Q8SR93	Q8SR93 encephalito
30	344	9.5	643	10	Q9LSL2	Q9LSL2 arabidopsis
31	339.5	9.3	461	10	Q9S7T1	Q9S7T1 arabidopsis
32	339	9.3	678	10	Q9LXA7	Q9LXA7 arabidopsis
33	337	9.3	602	10	Q9AR00	Q9AR00 lycopersico
34	323.5	8.9	766	10	Q9LNC4	Q9LNC4 arabidopsis
35	322.5	8.9	247	5	Q8T3Z6	Q8T3Z6 drosophila
36	318.5	8.8	268	5	Q9WIS2	Q9WIS2 drosophila
37	318	8.7	440	10	Q9LMU8	Q9LMU8 arabidopsis
38	315	8.7	487	10	Q93ZU2	Q93ZU2 arabidopsis
39	314	8.6	2429	11	Q8QZV8	Q8QZV8 mus musculu
40	303.5	8.3	818	10	Q9LKZ7	Q9LKZ7 arabidopsis
41	300.5	8.3	3080	5	Q9VRY3	Q9VRY3 drosophila
42	297.5	8.2	374	5	O76561	O76561 caenorhabdi
43	285	7.8	1188	10	Q41805	Q41805 zea mays (m
44	282.5	7.8	1235	4	Q9HOE9	Q9HOE9 homo sapien
45	282	7.7	369	10	Q9FT54	Q9FT54 arabidopsis

#### ALIGNMENTS

RESULT 1

O60433 PRELIMINARY; PRT; 731 AA.  
AC O60433  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE R31546\_1 (Fragment).  
GN HUNK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
RA Burkhart-Schultz K.J., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,  
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
RA Kobayashi A., Olsen A.S., Carrano A.V.;  
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RL "Sequence analysis of an ~1.5 MB OLF cluster in 19p13.1.;"  
RL EMBL: AC004798; AAC27978.1; -  
DR HSSP: Q92831; 1B91.  
DR InterPro: IPR001487; Bromodomain.  
DR Pfam: PF00439; bromodomain; 2.  
DR PRINTS: PR00503; BROMODOMAIN.  
DR SMART: SM00297; BROMO; 2.  
DR PROSITE: PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE: PS50014; BROMODOMAIN\_2; 2.  
FT NON\_TER  
SQ SEQUENCE 731 AA; 81383 MW; F56C3A02F1A26F65 CRC64;

Query Match 97.0%; Score 3531.5; DB 4; Length 731;  
Best Local Similarity 98.0%; Pred. No. 2.1e-200;  
Matches 672; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

```
QY 1 MSASGPGTRLNRLNLPVMDGLETSMSTTQAQAQOPQANAASTNPPPETSNPNKPKROT 60
DB 1 MSASGPGTRLNRLNLPVMDGLETSMSTTQAQAQOPQANAASTNPPPETSNPNKPKROT 71
QY 61 NOLQYLLRVVLKTLWKHGFAMFPQOPVDAVKLNLPDYKYIKTMDMGTIKKRLNNYIW 120
DB 72 NOLQYLLRVVLKTLWKHGFAMFPQOPVDAVKLNLPDYKYIKTMDMGTIKKRLNNYIW 131
QY 121 NAEICIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFQKINELPTEETEIMIVQAKGRG 180
DB 132 NAEICIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFQKINELPTEETEIMIVQAKGRG 191
QY 181 RGRKETGTAKEGVSIVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 240
DB 192 RGRKETGTAKEGVSIVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 251
QY 241 VYVPPQLOTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTTI 300
DB 252 VYVPPQLOTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTTI 311
QY 301 DPHIPEPPLPEPKTKTLGQRESSRPVKPKKDVPSQQHPAPEKSKVSEQLKCCSGI 360
DB 312 DPHIPEPPLPEPKTKTLGQRESSRPVKPKKDVPSQQHPAPEKSKVSEQLKCCSGI 371
QY 361 LKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 420
DB 372 LKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 431
QY 421 DVLRFMSNCYKYNPDHEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPPTKV 480
DB 432 DVLRFMSNCYKYNPDHEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPPTKV 491
QY 481 APPSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKKE 540
DB 492 APPSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKKE 551
QY 541 KDKKKEKKKKEEVENKSKAKEPPPKTKKNSSNSNSVSKKEPAPMKSPPPPYE 600
DB 552 KDKKKEKKKKEEVENKSKAKEPPPKTKKNSSNSNSVSKKEPAPMKSPPPPYE 611
QY 601 SEEDKCKPMSEYERKQLSLDINKLPGEKLGVRVHHIQSREPSLKNPNPDEIDFETLK 660
DB 612 SEEDKCKPMSEYERKQLSLDINKLPGEKLGVRVHHIQSREPSLKNPNPDEIDFETLK 671
QY 661 PSTLRELKALCHLLFAEEKETFKLRK 686
DB 672 PSTLREL-----ERYVTSCLRK 688
```

## RESULT 2

```
Q8VHF7 ID Q8VHF7 PRELIMINARY; PRT; 723 AA.
AC Q8VHF7;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Bromodomain-containing protein BRD4 short variant.
GN BRD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
[1]
RN RP SEQUENCE FROM N.A.
RA Houzelstein D., Bullock S.L., Lynch D.E., Grigorieva E.F.,
RA Wilson V.A., Biddington R.S.P.;
RT "Growth and early post implantation defects in mice mutant for the
RT bromodomain-containing protein Brd4."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF461396; AAL67834.1; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
```

```
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS0014; BROMODOMAIN_2; 2.
SQ SEQUENCE 723 AA; 80635 MW; 738F2AE5F58A56BC CRC64;
```

Query Match 93.3%; Score 3397; DB 11; Length 723;  
Best Local Similarity 94.3%; Pred. No. 1.7e-192;  
Matches 648; Conservative 5; Mismatches 24; Indels 10; Gaps 2;

```
QY 1 MSASGPGTRLNRLNLPVMDGLETSMSTTQAQAQOPQANAASTNPPPETSNPNKPKROT 60
DB 1 MSTSGPGRLNRLNLPVMDGLETSMSTTQAQAQOPQANAASTNPPPETSNPNKPKROT 60
QY 61 NOLQYLLRVVLKTLWKHGFAMFPQOPVDAVKLNLPDYKYIKTMDMGTIKKRLNNYIW 120
DB 61 NOLQYLLRVVLKTLWKHGFAMFPQOPVDAVKLNLPDYKYIKTMDMGTIKKRLNNYIW 120
QY 121 NAEICIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFQKINELPTEETEIMIVQAKGRG 180
DB 121 NAEICIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFQKINELPTEETEIMIVQAKGRG 180
QY 181 RGRKETGTAKEGVSIVPNTTQASTPQOTQOPN-PPVQATPHPPAVTDLIVQTPVM 239
DB 181 RGRKETGTAKEGVSIVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVM 240
QY 240 TVVPPQLOTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTT 299
DB 241 TVVPPQLOTPSPVPPQOPPPAPVQVQSHPPPIIAATPQVKTGKVKRKADTTPTT 300
QY 300 IDPHIPEPPLPEPKTKLGQRESSRPVKPKKDVPSQQHPAPEKSKVSEQLKCCSG 359
DB 301 IDPHIPEPPLPEPKTKLGQRESSRPVKPKKDVPSQQHPAPEKSKVSEQLKCCSG 360
QY 360 ILKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERF 419
DB 361 ILKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERF 420
QY 420 ADVRLMFNSCYKYNPDHEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPPTKV 475
DB 421 ADVRLMFNSCYKYNPDHEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPPTKV 480
QY 480 VAPSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKK 539
DB 481 VAPSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKK 540
QY 540 EKDKKKEKKKKEEVENKSKAKEPPPKTKKNSSNSNSVSKKEPAPMKSPPPPY 599
DB 541 EKDKKKEKKKKEEVENKSKAKEPPPKTKKNSSNSNSVSKKEPAPMKSPPPPY 600
QY 600 ESEEDKCKPMSEYERKQLSLDINKLPGEKLGVRVHHIQSREPSLKNPNPDEIDFETL 659
DB 601 ESEEDKCKPMSEYERKQLSLDINKLPGEKLGVRVHHIQSREPSLKNPNPDEIDFETL 660
QY 660 KPSLRELKALCHLLFAEEKETFKLRK 686
DB 661 KPSTLREL-----ERYVTSCLRK 678
```

## RESULT 3

```
Q9ESU6 ID Q9ESU6 PRELIMINARY; PRT; 1400 AA.
AC Q9ESU6;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Cell proliferation related protein CAP.
GN BRD4 OR CAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE-20396330; PubMed-10938129;
RA Dey A., Ellenberg J., Farina A., Coleman A.E., Maruyama T.,
RA Sciortino S., Lippincott-Schwartz J., Ozato K.;
RT "A bromodomain protein, MCPAP, associates with mitotic chromosomes and
RT affects G(2)-to-M transition."
RL Mol. Cell. Biol. 20:6537-6549(2000).
DR EMBL; AF273217; NAG02191.1; -.
DR HSSP; Q92831; 1891.
DR MGD; MGI:1888520; Brd4.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 1400 AA; 155923 MW; 9902BFF7B00ADB59 CRC64;

Query Match 93.2%; Score 3392; DB 11; Length 1400;
Best Local Similarity 94.2%; Pred. No. 6.8e-192;
Matches 647; Conservative 5; Mismatches 25; Indels 10; Gaps 2;

QY 1 MSAESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPNKPKROT 60
DB 1 MSTESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPNKPKROT 60

QY 61 NQLQYLLRVVLKTLWKHQFAWPPQOPVDVAVKLNLPDYKIIKTPMDMGTKIKRLNNYY 120
DB 61 NQLQYLLRVVLKTLWKHQFAWPPQOPVDVAVKLNLPDYKIIKTPMDMGTKIKRLNNYY 120

QY 121 NAQECIDQFNTMTNCTNYYNKPQDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
DB 121 NAQECIDQFNTMTNCTNYYNKPQDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180

QY 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
DB 181 RGRKETGAKPGVSTVNTTQASTPQTPQPNPPPPVQATPHPPFAVTPDLIAQPPVM 240

QY 240 TVVPPQPLQTPSPVPQPPAPQPVQSHPPPIAATPQVKTGKVRKADTTPTT 299
DB 241 TMVPPQPLQTPSPVPQPPAPQPVQSHPPPIAATPQVKTGKVRKADTTPTT 300

QY 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVDSQOHPAPEKSKVSSEOLKCCSG 359
DB 301 IDPIHEPPSLAPEPKTKLQORRESSRPVKPKDVDSQOHPEKSKSISEOLKCCSG 360

QY 360 ILKEMFAKKAAYAWPPYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
DB 361 ILKEMFAKKAAYAWPPYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 420

QY 420 ADVRLMFSNCKYKYNPDHEVAVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
DB 421 ADVRLMFSNCKYKYNPDHEVAVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 480

QY 480 VAPSSSDSSSDSSSDSDTDEERAQRLAEQLKAVHEQLAALSOQONKPKKK 539
DB 481 VAPSSSDSSSDSSSDSDTDEERAQRLAEQLKAVHEQLAALSOQONKPKKK 540

QY 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
DB 541 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 600

QY 600 ESEEDCKPMSYEKQLSLIDNKLPGELGRVHHIIQSRPSLKNPNDEIDEFTL 659
DB 601 ESEEDCKPMSYEKQLSLIDNKLPGELGRVHHIIQSRPSLKNPNDEIDEFTL 660

QY 660 KPSTLRELGAHLALFAEKETFKLR 686
DB 661 KPSTLREL-----ERVVTSCLR 678
```

```
RESULT 4
Q8VHF8 PRELIMINARY; PRT; 1400 AA.
ID Q8VHF8
AC Q8VHF8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bromodomain-containing protein BRD4 long variant.
GN BRD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Houzelstein D., Bullock S.L., Lynch D.E., Grigorieva E.F.,
RA Wilson V.A., Beddington R.S.P.;
RT "Growth and early post implantation defects in mice mutant for the
RT bromodomain-containing protein Brd4."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF461395; AAL67833.1; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 1400 AA; 155925 MW; 89952B9E75501BC4 CRC64;

Query Match 93.2%; Score 3392; DB 11; Length 1400;
Best Local Similarity 94.2%; Pred. No. 6.8e-192;
Matches 647; Conservative 5; Mismatches 25; Indels 10; Gaps 2;

QY 1 MSAESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPNKPKROT 60
DB 1 MSTESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPNKPKROT 60

QY 61 NQLQYLLRVVLKTLWKHQFAWPPQOPVDVAVKLNLPDYKIIKTPMDMGTKIKRLNNYY 120
DB 61 NQLQYLLRVVLKTLWKHQFAWPPQOPVDVAVKLNLPDYKIIKTPMDMGTKIKRLNNYY 120

QY 121 NAQECIDQFNTMTNCTNYYNKPQDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
DB 121 NAQECIDQFNTMTNCTNYYNKPQDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180

QY 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
DB 181 RGRKETGAKPGVSTVNTTQASTPQTPQPNPPPPVQATPHPPFAVTPDLIAQPPVM 240

QY 240 TVVPPQPLQTPSPVPQPPAPQPVQSHPPPIAATPQVKTGKVRKADTTPTT 299
DB 241 TMVPPQPLQTPSPVPQPPAPQPVQSHPPPIAATPQVKTGKVRKADTTPTT 300

QY 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVDSQOHPAPEKSKVSSEOLKCCSG 359
DB 301 IDPIHEPPSLAPEPKTKLQORRESSRPVKPKDVDSQOHPEKSKSISEOLKCCSG 360

QY 360 ILKEMFAKKAAYAWPPYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
DB 361 ILKEMFAKKAAYAWPPYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 420

QY 420 ADVRLMFSNCKYKYNPDHEVAVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
DB 421 ADVRLMFSNCKYKYNPDHEVAVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 480

QY 480 VAPSSSDSSSDSSSDSDTDEERAQRLAEQLKAVHEQLAALSOQONKPKKK 539
DB 481 VAPSSSDSSSDSSSDSDTDEERAQRLAEQLKAVHEQLAALSOQONKPKKK 540

QY 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
DB 541 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 600
```

```
QY 600 ESEEDCKPMSYBEKRLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETL 659
|||||
DB 601 ESEEDCKPMSYBEKRLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETL 660
|||||
QY 660 KPSTRELKALCHLLFAEKEFKLRK 686
|||||
DB 661 KPSTREL-----ERYVTSLRK 678
|||||

RESULT 5
OSJ125 PRELIMINARY; PRT; 726 AA.
ID Q9J125
AC Q9J125
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Bromodomain-containing FSH-like protein FSRG2.
GN BRD3 OR 2410084F24RIK OR FSRG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Shang E., Wolgemuth D.J.;
RT "Cloning and expression pattern of Fsr2, a putative murine
RT bromodomain-containing homolog of the Drosophila gene female sterile
RT homeotic.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF269193; AAF78072.1; -.
DR HSSP; Q92831; I891.
DR MGD; MGI:1914632; Brd3.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 726 AA; 79745 MW; 7AB3BADAD38A78F4 CRC64;

Query Match 52.6%; Score 1915; DB 11; Length 726;
Best Local Similarity 60.2%; Pred. No. 3.6e-105;
Matches 401; Conservative 64; Mismatches 141; Indels 60; Gaps 16;

QY 26 MSTTQAQAPQAPANAASNPPTTSNPKPKRTNQLQYLLRVLTKLWKHQPAPFQQ 85
|||||
DB 1 MSTTAAATGTPAVPGVNPPTPEVSNSKPGKTKNQLQYQWQVVKLWKHQPAPFQQ 60
|||||
QY 86 PVDAYKLNLPDYKIITPTMDGTIKRLNNYNNYAQECIQDFTMTNCTYIYNKPGDD 145
|||||
DB 61 PVDAILKLNLPDYHKIIPMDMGTIKRLNNYNNYWSASECQDFTMTNCTYIYNKPTDD 120
|||||
QY 146 IVLMAEALFKLQINELPTEETIMIVQAKGR-----GRKGTGAKGVSTVNTQTQ 201
|||||
DB 121 IVLMAEALFKLQINELPTEETIMIVQAKGR-----GRKGTGAKGVSTVNTQTQ 178
|||||
QY 202 ASTPPTQTPQPN-PPVQVATPHPPPAVTPDLIVQTVVTVVPPQPLQTPPPVPPQPP 260
|||||
DB 179 A-----TPQNIPPTVSQTP-----VIAATPVPTIT-----ANVTSPVPP 214
|||||
QY 261 PAPAPQVQSHPPPIAATPQPVKTKKGVKPKADTTPT-----TIDPIHEPSPLPPEKTT 316
|||||
DB 215 PAAPPPPATPIVPVPPTP-PVVKKGVKPKADTTPTTSAITASRSSESPPL-SEPKQA 272
|||||
QY 317 KLGQRRES-SRPVKPKDVPDSQHPAPEKSSKVSQOLKCCSGILKEMFAKHAAYAMP 375
|||||
DB 273 KVARESGGRPIKPKKDLDEGEVQHPAGKKGKGLSEHLRCHDSILREMLSKKHAAYAMP 332
|||||
QY 376 FYKPVVDVALGLHDYCDIIKHPMDMTIKSKLEAREYDAQEFQADVRLMSNCKYKNPP 435
|||||
```

```
DB 333 FYKPVDAEALHLDYHDIHKHPMDLSTVYKRMDSREYDPAQFAADIRLMSNCKYKNPP 392
|||||
QY 436 DHEVYVAMARKLQDVEMRFAMKMPDEPEPVPVAVSSPVPPTKVVAPSSSDSSSDSSD 495
|||||
DB 393 DHEVYVAMARKLQDVEMRFAMKMPDEPE-----APALPAPTAPIVSKGASRSSESS 446
|||||
QY 496 SDSSTDDSEERAOQLAEQLKAVHBOALALSOPOQNKP-KKKEKDKKKEK-----547
|||||
DB 447 SDGSSDSEERATRLAEQLKAVHBOALALSOPOQNKP-KKKEKDKKKEK-----506
|||||
QY 548 KEKHKKKEVEENKSK-----AKEPPPK--TKKNNSSNVSKKEPAPMKSKPPPTVES 601
|||||
DB 507 KEKEKHKAKSEEEKAKAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 564
|||||
QY 602 EEDCKPMSYBEKRLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETLKP 661
|||||
DB 565 EEEEGPLMSYDEKRLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETLKP 624
|||||
QY 662 STLREL 667
|||||
DB 625 TTLREL 630
|||||

RESULT 6
O88411 PRELIMINARY; PRT; 798 AA.
ID O88411
AC O88411;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Female sterile homeotic-related protein Frg-1.
GN BRD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhee K., Brunori M., Besset V., Wolgemuth D.J.;
RT "Expression and potential role of Frg-1, a putative murine
RT bromodomain-containing homolog of the Drosophila gene female sterile
RT homeotic.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045462; AAC24810.1; -.
DR HSSP; Q92831; I891.
DR MGD; MGI:99495; Brd2.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 798 AA; 88063 MW; A9942517CF15B7A1 CRC64;

Query Match 52.2%; Score 1901; DB 11; Length 798;
Best Local Similarity 54.9%; Pred. No. 2.7e-104;
Matches 403; Conservative 86; Mismatches 151; Indels 94; Gaps 20;

QY 5 SGPGTRLNLPVMDGLETSMSTTQAQAPQAPANAASNPPTTSNPKPKRTNQLQ 64
|||||
DB 26 AAGPKRIKPKSLLYEGFESPTMSVPA-----LQAPANPPPPVSNPKPKGRVNTQLQ 79
|||||
QY 65 YLLRVVLTALKHQPAPFQFPQVDAVKLNLPDYKIITPTMDGTIKRLNNYNNYAQEC 124
|||||
DB 80 YLHKVVMKALWKHQPAPFQFPQVDAVKLGLPDYHKIIPQMDGTIKRLNNYNNYAASE 139
|||||
QY 125 CIQDFTMTNCTYIYNKPGDDIVLMAEALFKLQINELPTEETIMIVQAKGR-----RG- 180
|||||
DB 140 CMQDFTMTNCTYIYNKPTDDIVLMAEALFKLQINELPTEETIMIVQAKGR----- 199
|||||
QY 181 -----RGKGTGAKGVSTVNTTQASTPTQPTQPPQPPVQATPFPFPAVTPDLIVQ 235
|||||
DB 200 KLAALQGSITSAHQVPVAVSSVSHYALYTPPEI-----PTVLNIPHP-----SVIS 246
|||||
```



```
QY 236 TPVMTVVPQIQTPTPPVPPQPPAPAPQPVQSHPPPIAATPQPVKTKGKVRKADTT 295
DB 247 SPLKSLH-----SAGPPLLAWSAAPP-----QPLAKKKGKVRKADTT 285
QY 296 TPTTIDPIHEP--PSLPP--EPKTKL--GORRESSRPVKPKKQVDPDSQOHPAPEKSSK 349
DB 286 TPTP-TAILAPGSPASPPGSLPEKAAARLPPMRRESGRPIKPPKRLDPDSQOQHSSKKKG 344
QY 350 VSEQLKCCSGILKEMFAKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDMSTIKSLEA 409
DB 345 LSEQLKHCNGILKELLSKKAHAAYAWPFYKPDVVEALGLHDYHDIHKHPMDLSTVKRKMEN 404
QY 410 REYRDAQEFAGADVRLMFSNCYKYNPPDHVEMVAMARKLDQVFMFAKMPDEPEEPVAVS 469
DB 405 ROYRDAQEFAGADVRLMFSNCYKYNPPDHVEMVAMARKLDQVFMFAKMPDEPEEPVAVS 464
QY 470 SPAVPPP-TKVVAPPSSSDSSSS-----SDSDSTDDSEERAOHLAEQ 517
DB 465 STALPPLTKSSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESS 524
QY 518 LKAVHEQLAALSOPOONKPKKKEKKKK-----EKHKRKEVEENKKS-KAKEPP-PK 571
DB 525 LRAVHEQLAALSOQPIKPKRK-REKKEKKRKAKEHGRIGIDEDDKGPRAPRPPQPK 583
QY 572 KTKKNSSNSNV-----SKKEPAPMKSKPPPT-----YESEEDCKCKPMSE 613
DB 584 KSKKAGGGSNATTLSHPGFGTSGGSNKLKPKKSOKTAPVLPVLTGYDSEESRPMSE 643
QY 614 EKROLSLDKLPGKGLGRVWHIIQSRPSLKNPNDETEIDFETLKPSTLRELGA-LCH 672
DB 644 EKROLSLDKLPGKGLGRVWHIIQAREPSLRNPNPEEIDEIDFETLKPSTLRELGVLS 703
QY 673 LLFAEEKETFKLRK 686
DB 704 CLRKKPKPYTIRK 717

RESULT 7
OS4795 PRELIMINARY; PRT; 798 AA.
AC OS4795;
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE RING3 protein.
GN BRD2 OR FSRG1 OR RING3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Thorpe K.L., Beck S.;
RL Immunogenetics 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA Taniguchi Y., Matsuzaka Y., Fujimoto H., Miyado K., Kohda A.,
RA Okumura K., Kimura M., Inoko H.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 47-549 FROM N.A.
RC STRAIN=INBRED CD-1; TISSUE=TESTIS;
RA Taniguchi Y., Matsuzaka Y., Fujimoto H., Miyado K., Kohda A.,
RA Okumura K., Kimura M., Inoko H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT *Sequence of the mouse major histocompatibility locus class II
```

```
RT region.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL009226; CAA15818.1; -
DR EMBL; AL009226; CAA15819.1; -
DR EMBL; D89801; BAA25416.1; -
DR EMBL; AB010248; BAA24379.1; -
DR EMBL; AB010247; BAA24378.1; -
DR EMBL; AB010246; BAA24377.1; -
DR EMBL; AF100956; AAC69907.1; -
DR HSSP; Q92831; I891.
DR MGD; MGI:99495; Brd2.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS00633; BROMODOMAIN_2; 2.
DR PROSITE; PS00633; BROMODOMAIN_3; 2.
SQ SEQUENCE 798 AA; 88066 MW; 08DD57FBF1385E96 CRC64;

Query Match 52.2%; Score 1901; DB 11; Length 798;
Best Local Similarity 54.9%; Pred. No. 2.7e-104;
Matches 403; Conservative 86; Mismatches 151; Indels 94; Gaps 20;

QY 5 SGPGLRLNLPVMDGLETSONSTTOAQOPOQAPANAASNPPTSNPNKPKROTNIQ 64
DB 26 AAPGKIRKPSLYEGFESPTMASVPA-----LQLAPANPPPEVSNPKKGRVTNQLQ 79
QY 65 YLLRVKTKIMKHQFAWPFQPDVAVKLVNLYKIKTKPMDMGTIKKRLNNYNNAAOE 124
DB 80 YLHKVVMKALWKHQAQFQPDVAVKLVNLYKIKTKPMDMGTIKKRLNNYNNAAOE 139
QY 125 CIOQFNTMTNCTYIYNKPGDDIVLMAEALKEFLQKINELPTEETIMIVQAKG---RG- 180
DB 140 CMQDENTMTNCTYIYNKPGDDIVLMAEALKEFLQKINELPTEETIMIVQAKG 199
QY 181 -----RGRKETGTAKPGVSTVNTTQASTPPTQTPQPNPPVQATPHPPATVPLIVQ 235
DB 200 KLAALQGGSTSAHQVPAVSVSHTALYTPPPEI-----PTTVLNIHPH-----SVIS 246
QY 236 TPVMTVVPQIQTPTPPVPPQPPAPAPQPVQSHPPPIAATPQPVKTKGKVRKADTT 295
DB 247 SPLKSLH-----SAGPPLLAWSAAPP-----QPLAKKKGKVRKADTT 285
QY 296 TPTTIDPIHEP--PSLPP--EPKTKL--GORRESSRPVKPKKQVDPDSQOHPAPEKSSK 349
DB 286 TPTP-TAILAPGSPASPPGSLPEKAAARLPPMRRESGRPIKPPKRLDPDSQOQHSSKKKG 344
QY 350 VSEQLKCCSGILKEMFAKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDMSTIKSLEA 409
DB 345 LSEQLKHCNGILKELLSKKAHAAYAWPFYKPDVVEALGLHDYHDIHKHPMDLSTVKRKMEN 404
QY 410 REYRDAQEFAGADVRLMFSNCYKYNPPDHVEMVAMARKLDQVFMFAKMPDEPEEPVAVS 469
DB 405 ROYRDAQEFAGADVRLMFSNCYKYNPPDHVEMVAMARKLDQVFMFAKMPDEPEEPVAVS 464
QY 470 SPAVPPP-TKVVAPPSSSDSSSS-----SDSDSTDDSEERAOHLAEQ 517
DB 465 STALPPLTKSSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESS 524
QY 518 LKAVHEQLAALSOPOONKPKKKEKKKK-----EKHKRKEVEENKKS-KAKEPP-PK 571
DB 525 LRAVHEQLAALSOQPIKPKRK-REKKEKKRKAKEHGRIGIDEDDKGPRAPRPPQPK 583
QY 572 KTKKNSSNSNV-----SKKEPAPMKSKPPPT-----YESEEDCKCKPMSE 613
DB 584 KSKKAGGGSNATTLSHPGFGTSGGSNKLKPKKSOKTAPVLPVLTGYDSEESRPMSE 643
QY 614 EKROLSLDKLPGKGLGRVWHIIQSRPSLKNPNDETEIDFETLKPSTLRELGA-LCH 672
DB 644 EKROLSLDKLPGKGLGRVWHIIQAREPSLRNPNPEEIDEIDFETLKPSTLRELGVLS 703
QY 673 LLFAEEKETFKLRK 686
DB 704 CLRKKPKPYTIRK 717
```

[illegible]

```
Db 256 EAAKAIIPARRSGRPKPKKDLPSQOQQT-SKKGLSEQLKYNGILLKELLSSKHAA 314
QY 372 YAMFYKPVDEALGLHDYCDIIKHPMDSTIKSKLEAREYRDAQEGADVRLMFSNCYK 431
Db 315 YAMFYKPVDSALGLHDYHEIIKHPMDLSTIKRKMRNDYHDAQEAFADVRLMFSNCYK 374
QY 432 YNPDPHEVYAMARKLDQVFEMRFKAPMDEPEE-PVAVSSPAVPPPTKVAVPSSSD--- 487
Db 375 YNPDPHDVYAMARKLDQVFESVAKMPDEPDQASPPSVSAPLGLGALSSESSDEDD 434
QY 488 -----SSSDSSSDSDSDSEERAOALAEQLKAVHEQLAALSQPOONK 535
Db 435 EDEDEDDDESSSSSSSSSESS--DSEERANRLAEQLKAVHEQLAALSQGPVSK 493
QY 536 PKKKEKDKKKKKKKKEEVENKSKAKPPPKTKK-----NNSNSNVSKKKEPAPM 591
Db 494 PKKREKKKKKKKKKKGRGDEE--SRARQAQLRKKAKKGGGGGGSSKSKKAAGA 550
QY 592 KSKPPPT-YESEEDCKPKSYEEKQLSLDINKLPGEKLGVRVHIIQSREPSLKNSNP 650
Db 551 ALPPPTPLDYSEEESESKPMYDEKRLSLDINKLPGEKLGVRVHIIQSREPSLRNSPE 610
QY 651 EIEIDFETLKPSTLREL 667
Db 611 EIEIDFETLKPSTLREL 627

RESULT 10
Q8QFT7 PRELIMINARY; PRT; 701 AA.
AC Q8QFT7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DE 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Hypothetical 76.4 kDa protein.
GN BRD3
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21845882; PubMed=11856876;
RA Bouchireb N., Grutzner F., Haaf T., Stephens R.J., Elgar G.,
RA Green A.J., Clark M.S.;
RT "Comparative mapping of the human 9q34 region in Fugu rubripes.";
RL Cytogenet. Cell Genet. 94:173-179(2001).
DR EMBL; AJ311635; CAC84085.1;
KW Hypothetical protein.
SQ SEQUENCE 701 AA; 76356 MW; 74E1F328615E9BD0 CRC64;

Query Match 48.3%; Score 1757; DB 13; Length 701;
Best Local Similarity 56.4%; Pred. No. 7.1e-96;
Matches 372; Conservative 68; Mismatches 151; Indels 68; Gaps 15;

QY 26 MSTTQAOQPQANAASTNPPPETSNPKPKRQTQLOQLYLLRVLTKLWKHFQFWFPQ 85
Db 1 MSDAPEAAPSP--PPLTNPPPEVSNPNKPGKRTQLOLYMQNVVYVTKLWKHFQFWFPQ 58

QY 86 PVDVAVLNDYKYIKITPMDMGTIKRLENNYVNAQECIQDFNTFTNCYINPKGDD 145
Db 59 PVDAILCLADYHKVINKPNMDMTIKRLENNYVNAQECIQDFNTFTNCYINPKGDD 118

QY 146 IVLMAFALEKLFQKINELPTEETEMIVQAKGRCGRKETGTAKPCVSTPN---TTQA 202
Db 119 IVLMAQALEKIFQKVAQMPQEEVALLPPAPKGNKSK-----QPAAGTIVSQQAESSA 172

QY 203 STPTQOTQTPQNPVPPVQATPHPPFVATVDTLVQTPVMTVVPVPPQTPPTPPVPPQPPPA 262
Db 173 SSPSPSPSPSPSTPWTIST-----TPTPVQTPPI-SAPQPPAA 210
```

```
QY 263 PAPQVOSHPIITAATQPVKTKKGVKVKADTTPTT--IDPIHEPSPSPPEPKTKLG- 319
Db 211 MMP-----SAQPVVKKKGVKVKADTTPTTSAISAGRADSPSAQDAKPAKLG 258
QY 320 -QRRESSRPVKPKP---KQVDPDQOQHAPBKSSKVSQKCCSGILKEMFAKHAAYAWP 375
Db 259 TRREATAPARTKREVEEYAGGAGGGRKTKLGGQMKHCDAILKEMLSKHAAYAWP 318
QY 376 FYKPDVDEALGLHDYCDIIKHPMDSTIKSKLEAREYRDAQEGADVRLMFSNCYKYNPP 435
Db 319 FYKPDAAELHLDYHDIKHPMDLSTIRKMDKGEYNEPQSFATDVRLMFSNCYKYNPP 378
QY 436 DHEVYAMARKLDQVFEMRFKAPMDEPEEPVAVSSPAVPPPTKVAVPSSSDSSSDSS 495
Db 379 DHEVYAMARKLDQVFEMRFKAPMDEPEEPVAVSSPAVPPPTKVAVPSSSDSSSDSS 432
QY 496 SDSTSDSEERAOALAEQLKAVHEQLAALSQPOONKPK-KKKEKDKKKKKK-KHR 553
Db 433 ESS--DSEERATRLAEQLKAVHEQLAVLSQAPVSKPKKKKKKKKKKKKGNKA 489
QY 554 KEEVEENKSKAKPPPKTKK--NNSNSNVSKKEPAPMKSKPP---PTVESEEDCKK 608
Db 490 KMEEEKPKTTAQPKPANOKKPPARKANSTVTATROPKSSKTSVSGSTNGDDGESAL 549
QY 609 PMSYEKRLSLDINKLPGEKLGVRVHIIQSREPSLKNSNPDEIEIDFETLKPSTLREL 667
Db 550 PMSYDEKRLSLDINKLPGEKLGVRVHIIQTRPSLRSDNPDEIEIDFETLKPSTLREL 608

RESULT 11
Q8UUM2 PRELIMINARY; PRT; 814 AA.
AC Q8UUM2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RING3 protein.
GN RING3.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD-RR;
RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;
RT "Nucleotide Sequence of the MHC Class I Region of a Teleost, the
RT Medaka.";
RL Immunogenetics 0:0-0(2002).
DR EMBL; AB073376; BAB83842.1;
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
SQ SEQUENCE 814 AA; 90116 MW; C71293789354623D CRC64;

Query Match 46.5%; Score 1690.5; DB 13; Length 814;
Best Local Similarity 49.6%; Pred. No. 6.9e-92;
Matches 377; Conservative 85; Mismatches 135; Indels 163; Gaps 21;

QY 1 MSAESGPGTBLRNLVPMGDGLETSMSTTQAOQPQANAASTNPPPETSNPKPKRQT 60
Db 24 MMDQSAAGKIRKPSLLYEDFESPGM-----PAISQMAPSGPPQPPVRDPSRHRMT 75

QY 61 NOLQYLRRVVLTKLWKHFQFWFPVATVDTLVQTPVMTVVPVPPQTPPTPPVPPQPPPA 120
Db 173 SSPSPSPSPSPSTPWTIST-----TPTPVQTPPI-SAPQPPAA 210
```

```
Db 76 NOLQFLQKVLKSLWRHFAWPEHPVDAAKLNLPDYHKIIKTPMDMGTIRKLENNYR 135
Qy 121 NQAECIQDNFTNFCYIYNKPGDDIVLMAEALFKLQKINELPTETETIMVOAKGRG 180
Db 136 SASCEQDNFTNFCYIYNKPGDDIVLMAEALFKLQKINELPTETETIMVOAKGRG 193
Qy 181 RGRKETSQAKPGVST--VPNTQA-STPQOTQPPQPPVQATPHPPPAVTPDLIVQT 236
Db 194 KOSKISG---GVTQAHQVAVSQSVSP--TPE-----TPDSILST 230
Qy 237 PMVTV-----VPPQLOTPPPVPOPPAPAPAPVQSHPPPIAATPPQVKTGKVKR 290
Db 231 PPQIILTKCSITLQPEQT---IPAITCPPT-----QPTAKKGVKR 270
Qy 291 KADTTPTTID-PI-----HEPP----- 307
Db 271 KADTTPTTAMPIMSTLGVGTIGLGMVGGHDSPLTLTSLGMDHSSGLMNOAISISQM 330
Qy 308 -----SLPEPKTKLQGRRESRPVKPKDVPDSQQHAPAEKSKVSEQLKCCSGI 360
Db 331 GGMGGGTMLGAKATSGSRGVSGRPIKPPKDLDPDSIL-PTPVRRSKLSPQLRYCSGV 389
Qy 361 LKEMFAKHAAYAMPFYKPDVVEALGHDYCDIHKPHMDSTIKSKLEAREYRDAQERGA 420
Db 390 LKELLSKHAAYAMPFYKPDVAVSLGLHDYHEIILKFPMDLSTIKRMDGREYRDAQFSA 449
Qy 421 DVLMFNSCYKYNPPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPTKV 480
Db 450 DVLMFNSCYKYNPPDHDVAMARKLDQVFECFAKMPDE-----APAPSSST--- 497
Qy 481 APPSSSSSSSSSSSSSSST-----DSEERARLAELQKAVHEQLAALQS 530
Db 498 -----SSSSSSSSSESSSSSSSESSSSSSSSSSSSSSSSSSSSSSSSSSSS 553
Qy 531 PQONKPKK--EKDKERK--KKHKKEE-----VEENKSKAKPPPKKT 573
Db 554 GPIVKPKKKKKKKKKRKEKRRRIEDDLTPIRPKAPKTKTKTRMDCPVIPM 613
Qy 574 KKNSSNNSVKKKPPAPKSKPP-----PYESEEEDCKPMSYEKRLQSLDINKLPG 627
Db 614 KKTOSKNSKSKSKAAITFNPHHDPLVGHFDSDEEDTAPMSYDEKRLQSLDINKLPG 673
Qy 628 EKLGRVHVHIIOSRPSLKNSPDEIDEIDFETLKPSLREL 667
Db 674 EKLGRVHVHIIOSRPSLKNSPDEIDEIDFETLKPSLREL 713

RESULT 12
Q8T775 PRELIMINARY; PRT; 664 AA.
AC Q8T775;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 74.6 kDa protein (Fragment)
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11967531.
RA Abi-Rached L., Gilles A., Shlina T., Pontarotti P., Inoko H.;
RT "Evidence of en bloc duplication in vertebrate genomes."
RL Nat. Genet. 0:0-0(2002).
DR EMBL; AF391288; AAM18869.1; -.
KW Hypothetical protein.
FT NON_TER 664 664
SQ SEQUENCE 664 AA; 74627 MW; BF3912BA045253BF CRC64;

Query Match 46.28; Score 1680; DB 5; Length 664;
Best Local Similarity 50.99; Pred. No. 2.3e-91;
Matches 361; Conservative 87; Mismatches 133; Indels 128; Gaps 25;
```

```
Qy 1 MSASGPGTTRLNLPVM--GDGLETSMSTTQAQPOQAPANAASNTNPPPTSNPNKPKRQ 59
Db 1 MNSPTGAGKRRRPSIMYAEGETSE--SKQA-----PKKPGRM 37
Qy 60 TNLOQLLRVYLKTLWKHQFAWFPQOPVDVAKLNLDPYKIIKTPMDMGTIKKRLNNY 119
Db 38 TNLOQLLRVYLKTLWKHQFAWFPQOPVDVAKLNLDPYKIIKTPMDMGTIKKRLNNY 97
Qy 120 WNAECIQDNFTNFCYIYNKPGDDIVLMAEALFKLQKINELPTETETIM-----IV 174
Db 98 YSAECIQDNFTNFCYIYNKPGDDIVLMAEALFKLQKINELPTETETIM-----IV 157
Qy 175 QAKGRGRKGTAKPGVSTVPNTQA-STPQOTQPPQPPVQATPHPPPAVTPDLI 233
Db 158 QTKGKKGPRK-----PRVQAAS--TAVPVNAQPPVQSPPPPAVTPPPI 203
Qy 234 VQT-----PMTVVPQPL--QTPPPV-----PP-----QPQPPAPAPQVQSH 271
Db 204 VPTTVAPTHVPTDLSLPQIPHQPVPVQPPVQPPPTTRAPPARTQQPRPPPIEQPQLQMP 263
Qy 272 P-----PTIAATPO--PVKTKGVKRRADTTPTTIDPIHEPPSLP--PEPKT 315
Db 264 PQLPQGSTAPAPLISQPGQAVTKVKKGKRRADTTPTTIDPIHEPPSLP--PEPKT 323
Qy 316 TKLGO--RESSRPVKKKQVDPDSQHPAPEKSKVSQLKCCSGILKEMFAKHAAYAW 374
Db 324 AKIPVRESGRQIKPPRRPELPETEQH--SSKKGKLSAQLKYCOGIIKEMFAKHAAYAW 382
Qy 375 PFYKPDVVEALGHDYCDIHKPHMDSTIKSKLEAREYRDAQERGAQVRLMFSNCYKYNP 434
Db 383 PFYKPDVVEALGHDYHEIILKPHMDSTIKSKLEAREYRDAQERGAQVRLMFSNCYKYNP 442
Qy 435 PDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSP---AVPPPTKVAP--PSSSDSSS 490
Db 443 PEHDVQMARKLDQVFEMFAKMPDEPDP-----EPQPLVPPPPAAPEPNTSTSSSS 497
Qy 491 DSSSDSSSTDDSEERARLAELQKAVHEQLAALQS--PQONKPK--KKKEKDK---K 544
Db 498 DSSSDSSST--DSEERARLAELQKAVHEQLAALQS--PQONKPK--KKKEKDK---K 555
Qy 545 EKKEKHKRKEEVEENKSKAKE-----PPPKTKKKNSS 579
Db 556 DKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 615
Qy 580 NSNVSKKEPAPKSK--PPPTYESEEEDCKPMSYEKRLQSLDINKLPG 626
Db 616 KTNSSRQSSKSKAQTTPAPAYESEDLDCKPMTYDEKRLQSLDINKLPG 664

RESULT 13
Q9GU61 PRELIMINARY; PRT; 732 AA.
AC Q9GU61;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RING3 (Fragment).
GN RING3.
OS Myxine glutinosa (Atlantic hagfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
OX Myxiniidae; Myxiniinae; Myxine.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP-ACTIVATED PERIPHERAL BLOOD LEUKOCYTES;
RA White G.P., Cunningham C.;
RT "Characterization of RING3-like protein from Atlantic hagfish (Myxine glutinosa).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF191032; AAG179.1; -.
DR HSP; 092831; 1891.
DR InterPro; IPR001487; Bromodomain.
```

```
DR Pfam: PF00439; bromodomain_2.  
DR PRINTS: PR00503; bromodomain.  
DR SMART: SM00297; BROMO; 2.  
DR PROSITE: PS00633; BROMODOMAIN_1; UNKNOWN_1.  
DR PROSITE: PS00014; BROMODOMAIN_2; 2.  
FT NON_TER 732 732  
SQ SEQUENCE 732 AA; 81669 MW; 2038D61BD20F6B50 CRC64;  
  
Query Match 46.1%; Score 1678; DB 5; Length 732;  
Best Local Similarity 49.5%; Pred. No. 3.4e-91;  
Matches 376; Conservative 71; Mismatches 165; Indels 148; Gaps 19;  
  
QY 26 MSTTQAAQAPQANAASTNPPPTSPNPKKROTNQLOYLRLVVLATLKKHQFAWPFQ 85  
DB 1 MATSPQOQQO---QPLCNPMPPPTNPKPRGLTNQLQFLQKVMKVKHQFAWPFHH 57  
  
QY 86 PVDVAVKLNLPDYKIIKTPMDMGTIKRLNNYWNNAQECIQDFTMTNCTYIYNKPGDD 145  
DB 58 PVDAAKLNLPDYQIKNPLDMLTIKRLNNYWNNAQECIQDFTMTNCTYIYNRND 117  
  
QY 146 IVLMAEALEKLFQKINELPTEETEMIVQAK-GRGRGKETGTAKPGVSTVPTNTQAST 204  
DB 118 IVLMAOTVEKAFQKVAEMPEVEEITSPVARVPQRRGRK-----PAALTAAG 165  
  
QY 205 PPQTQTPQNPVPVQATPHPP-----AVTPDLIVTPVMTVVPVQPLQ----- 248  
DB 166 PVVTAGVPVLSPIPAAPSPSQRPRAAAAATQOIAAAOTLPASPLRLLEQCGQPVQNM 225  
  
QY 249 -----TPPPV--PPQOPPPAPAPQPVQSHPIIAATP----- 279  
DB 226 ARSVSTPPAVGQPTLETPSAAGVQSLLSQPMVSLTPLNMQGFAATHTPTPKIDALGR 285  
  
QY 280 -----QPVTKKGVKVKKADTTTPT 298  
DB 286 AERQGEEDTAMEQSELAFKGGMRSGEINPDISISWVHTQENQASRGVKKKADTTTPT 345  
  
QY 299 TIDPIHPPSLPPPKTKLG-----QRESSRPVKPKKQVDSQO-HPA 343  
DB 346 AL-----QAPRNEETSPGLADGPAGIVRPLVPGRESSRPVKARKEVPDSPALPV 399  
  
QY 344 PEKSSKVSQKCCGILKEMFAKHAAYAWPFYKPYDVEALGHDYCDIHKHPMDMTI 403  
DB 400 SKRVQMSDQLRHQOTILKEIFTKHAAYAWPFYKAVDAFALGLHDYHDIKIPMDLTI 459  
  
QY 404 KSKLEAREYDAQFEGADVRLMFSNCKYKPNPDHEVVMARKLQDVFMFRFAPKMPDE--- 460  
DB 460 KERFEREYNLHEFADDMRLMFSNCKYKPNPDHEVVMARKLQDVFMFRFAPKMPDETPP 519  
  
QY 461 -PEPVPVAVSSPAVPPPTKVYAPP-----SSSDSSSDSSSDSDSDSDSEERAQRLAELQ 515  
DB 520 PPVQPPVTPQPTTTPP-----PPDVHLTSSSSSSSSSTDESS-DDSEKERANRLAELQ 574  
  
QY 516 EQLKAVHEQLAALSQPOONKPKKEDKKEKKK--HKRKEEVEENKSKAKEPPPKKT 573  
DB 575 EQLKALHEQLASLSQAIPSKPKK-KERKEKDKVLEKARKWEDEKVLKNLLQAKM 633  
  
QY 574 KKNSSNSNYSKKE-----PAPKSKP--PPYVESEEDCKPMSYEERKQLSLDINKL 625  
DB 634 GKQNSNRPIKKEBDSITGASSLNAIPNVPVTLPGDTEG-ARPMYDEKRLQSLDINRL 692  
  
QY 626 PGEKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 665  
DB 693 PGDKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 732  
  
RESULT 14  
Q91Y44 PRELIMINARY; PRT; 956 AA.  
AC Q91Y44;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Bromodomain-containing female sterile homeotic-like protein.
```

```
GN FSRG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Shang E., Wolgemuth D.J.;  
RT "The Murine Fstg (Female Sterile Homeotic-Related Gene) Family  
RT Consists of at Least Four Members Whose Restricted Patterns of  
RT Expression Suggest Functions During Spermatogenesis";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF358660; AAK50736.1; -;  
DR MGD; MGI:1891374; Fsr3.  
DR InterPro; IPR001487; Bromodomain.  
DR Pfam: PF00439; bromodomain; 2.  
DR PROSITE: PS00633; BROMODOMAIN_1; UNKNOWN_2.  
DR PROSITE: PS00014; BROMODOMAIN_2; 2.  
SQ SEQUENCE 956 AA; 107338 MW; 8474CAEDC6217AC6 CRC64;  
  
Query Match 42.3%; Score 1540; DB 11; Length 956;  
Best Local Similarity 51.8%; Pred. No. 5.9e-83;  
Matches 332; Conservative 71; Mismatches 138; Indels 100; Gaps 14;  
  
QY 41 ASTNPPPTSPNPKKROTNQLOYLRLVVLKTHQFAWPFQPVDAVKLNLPDYKI 100  
DB 9 AIVNPPPEYINTKSGRLTNQLQFLQRVVLKALWKHGSWPFQPVDAVKLNLPDYTI 68  
  
QY 101 IKTPMDMGTIKRLNNYWNNAQECIQDFTMTNCTYIYNKPGDDIVLMAEALEKLFQK 160  
DB 69 IKTPMDLNTIKRLNNYWNNAQECIQDFTMTNCTYIYNKPGDDIVVMAEALEKLFQK 128  
  
QY 161 INELPTEETEMIVQAKGRGRKETGTAKPGVSTVPTNTQASTPPTQTPQPNPPVQA 220  
DB 129 LSQMPQEEVQV-----GKERIKDI-QOKIATVSSAKEQI-----PSKA 165  
  
QY 221 TPHPF-----PAVTPDLIVTPVMTVVPVQPLQTPPVPPQPPAPAPQVQSHPII 275  
DB 167 AENFKRQEIPLSGPLDISL-----SPLNMAQEAPP-----I 197  
  
QY 276 AATPQPVKTKGVKVKADTTTPTI---DPHPPSPPEPKTKLGGRESSRPVKPK 332  
DB 198 CDSQSLVQITGVKRRADTTTPTIXIAKASSESPITL-----RETQPNVPV 244  
  
QY 333 KD-----VPDSQHPAPEKSSKVSQKCCGILKEMFAKHAAYAWPFYKPYDVEALG 386  
DB 245 KENTVKNVLPDSQGHKVLTKVTEQLKHCSEILKEMAKKHLPYAWPFYNPVDADALG 304  
  
QY 387 LHDYCDIHKHPMDMTSKLEAREYDAQFEGADVRLMFSNCKYKPNPDHEVVMARKL 446  
DB 305 LHNYDYVKNPMDLGTIKGMDNOEYKDAVEFAADVRLMFMNCKYKPNPDHEVVMARKL 364  
  
QY 447 QDVFMFRFAPKMPDEPEPVPVAVSSPAVPPPTKVYAPPSSSDSSSDSDSDSDSEEE 506  
DB 365 QDVFMFRFAPKMPDEPIESMHACHL-----TNSAQLSRESSEAS-GDASSEDE 417  
  
QY 507 RAQRLAELQKAVHEQLAALSQPOONKPKKEDKKEKKK--HKRKEEVEENKSKAKE 566  
DB 418 RVQHLAKLQQLNNAHQQLVLSQVPLRLKK--KNEKSRAPKPKVNNRDEPRKK-- 473  
  
QY 567 EPPPKTKKNSSNSNYSKKEPKAPKSKPPPTVESEEDCKPMSYEERKQLSLDINKLP 626  
DB 474 ---PKMQKQEKAKINQPKK-----KPLKSEEDNAKPMYDEKRLQSLDINKLP 522  
  
QY 627 GEKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 667  
DB 523 GDLGRIVHIIQSRPSLKNPNDEIDFETLKPSTLR 563  
  
RESULT 15  
O14789 PRELIMINARY; PRT; 947 AA.  
ID O14789
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 10:57:44 ; Search time 6007 seconds

(without alignments)  
11283.575 Million cell updates/sec

Title: US-09-700-590A-101

Perfect score: 2329  
Sequence: 1 gcaaggatactagcatgctc.....ngcctttaaagtlnccca 2329Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

GeneBank: 1: gb\_da: 2: gb\_da: 3: gb\_da: 4: gb\_da: 5: gb\_da: 6: gb\_da: 7: gb\_da: 8: gb\_da: 9: gb\_da: 10: gb\_da: 11: gb\_da: 12: gb\_da: 13: gb\_da: 14: gb\_da: 15: gb\_da: 16: gb\_da: 17: gb\_da: 18: gb\_da: 19: gb\_da: 20: gb\_da: 21: gb\_da: 22: gb\_da: 23: gb\_da: 24: gb\_da: 25: gb\_da: 26: gb\_da: 27: gb\_da: 28: gb\_da: 29: gb\_da: 30: gb\_da: 31: gb\_da: 32: gb\_da: 33: gb\_da: 34: gb\_da: 35: gb\_da: 36: gb\_da: 37: gb\_da: 38: gb\_da: 39: gb\_da: 40: gb\_da: 41: gb\_da:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2065	88.7	3149	9 HSHUNKI	Y12059 H. sapiens H
2	2065	88.7	3149	9 HSHUNKI	AF386649 Homo sapi
3	1724	74.0	2727	10 AF461396	AF461396 Mus muscu
4	1720.8	73.9	5281	10 AF273217	AF273217 Mus muscu
5	1256.2	53.9	2199	10 BC008532	BC008532 Mus muscu
6	663.2	28.5	3028	10 HUMOREX	D26362 Human mRNA
7	644.4	27.7	3033	10 BC031536	BC031536 Mus muscu
8	642.8	27.6	2933	10 AF269193	AF269193 Mus muscu
9	559.4	24.0	667	6 AX427508	AX427508 Sequence
10	541.6	23.3	2548	9 BC032124	BC032124 Homo sapi
11	461	19.8	3649	10 AF045462	AF045462 Mus muscu
12	460.2	19.8	4053	9 HSF5MR	X62083 H. sapiens m
13	460.2	19.8	4053	9 HUMESHG	M80613 Human homol
14	460.2	19.8	4053	9 HUMESHG	AX409010 Sequence
15	460.2	19.8	4664	6 AX409010	D42040 Human mRNA
16	460.2	19.8	4664	9 HUMKIA9001	AB010248 Mus muscu
17	403.2	17.3	1554	10 AB010248	AB010248 Mus muscu
18	403.2	17.3	1657	10 AB010247	AB010247 Mus muscu
19	403.2	17.3	1688	10 AB010246	AB010246 Mus muscu
20	398.4	17.1	2206	10 AF318183	AF318183 Mus muscu
21	393.4	16.9	3391	10 AF358660	AF358660 Mus muscu
22	383.4	16.5	3429	9 BC011898	BC011898 Homo sapi
23	381.6	16.4	2350	9 AK000477	AK000477 Homo sapi
24	371	15.9	1884	9 AK024312	AK024312 Homo sapi
25	367	15.7	42497	9 AC004798	AC004798 Homo sapi
26	366.2	15.7	4667	9 HSM804033	AL837222 Homo sapi
27	336.6	14.5	2532	5 AF191032	AF191032 Myxine gl
28	311.2	13.4	177465	2 AC090975	AC090975 Mus muscu
29	311.2	13.4	235218	2 AC074208	AC074208 Mus muscu
30	298	12.8	40649	9 AC003111	AC003111 Human DNA
31	294.2	12.6	3106	9 AF019085	AF019085 Homo sapi
32	275.6	11.8	207585	2 AC073779	AC073779 Mus muscu
33	255.6	11.0	236659	2 AL808021	AL808021 Mus muscu
34	253.6	10.9	3404	9 AK056504	AK056504 Homo sapi
35	208.2	8.9	1741	9 S78771	S78771 NAT-Cpg 1st
36	207	8.9	352	6 AX071305	AX071305 Sequence
37	203.8	8.8	5632	3 DROFSA	M23222 D. melanogas
38	203.8	8.8	7792	3 DROFSA	M23221 D. melanogas
39	203.8	8.8	37518	2 AC014003	AC014003 Drosophila
40	203.8	8.8	181178	3 AC023713	AC023713 Drosophila
41	203.8	8.8	306267	3 AE003442	AE003442 Drosophila
42	197.6	8.5	384	10 MMPJ32FSH	X97573 M. musculus
43	170.2	7.3	4651	5 GGRING3GE	X96669 G. gallus RI
44	170.2	7.3	37644	6 A95302	A95302 Sequence 33
45	170.2	7.3	92863	12 GGBLOCUS	AL023516 Gallus ga

## ALIGNMENTS

RESULT 1  
HSHUNKI  
LOCUS HSHUNKI 3149 bp mRNA linear PRI 02-MAY-1998  
DEFINITION H. sapiens HSHUNKI mRNA.  
ACCESSION Y12059  
VERSION Y12059.1 GI:3115203  
KEYWORDS HSHUNKI gene.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Weber, B.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3149)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Weber, B.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1997) B. Weber, Labor Paediatriche  
Molekularbiologie, Universitaetsklinikum Charite, Ziegelstr. 5-9,  
10098 Berlin, FRG

FEATURES Location/Qualifiers  
source 1..3149

gene /organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="placenta"  
223..2391  
/gene="HUNK1"  
223..2391  
/note="strong homology to human RING3 sequence"  
/protein\_id="CAA72780.1"  
/db\_xref="GI:3115204"  
CDS /db\_xref="SPTREMBL:O60885"

BASE COUNT 818 a 946 c 788 g 597 t  
ORIGIN  
4 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 63  
211 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 270  
64 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAACACAGGCCAGGCCCAACCCAG 123  
271 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAACACAGGCCAGGCCCAACCCAG 330  
124 CCAGCCAAAGCAGCAGCAGCAACCCCGCCCGCCGAGAGACCTCCCAACCTTAACAGCCC 183  
331 CCAGCCAAAGCAGCAGCAGCAACCCCGCCCGCCGAGAGACCTCCCAACCTTAACAGCCC 390  
184 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 243  
391 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 450  
244 CACCAAGTTGCAATGGCTTTCCAGCAGCCTGTGATGCCCTCAAGCTGAACTCCTGAT 303  
451 CACCAAGTTGCAATGGCTTTCCAGCAGCCTGTGATGCCCTCAAGCTGAACTCCTGAT 510  
304 TACTTAAGATCAATTAAGAGCCTATGATGGGAACAATAAGAGAGGCTTGGAAAAC 363  
511 TACTTAAGATCAATTAAGAGCCTATGATGGGAACAATAAGAGAGGCTTGGAAAAC 570  
364 AACTATTACGAGATGCTCAGGATGATATCCAGACTTCAACACTATGTTTCAAAATGT 423  
571 AACTATTACGAGATGCTCAGGATGATATCCAGACTTCAACACTATGTTTCAAAATGT 630  
424 TACATCTACACAAAGCCTGAGAGATGATGTTTAATGGCAGAACTCTGAAAAAGCTC 483  
631 TACATCTACACAAAGCCTGAGAGATGATGTTTAATGGCAGAACTCTGAAAAAGCTC 690  
484 TTCTTGGCAAAAAATAATAGCTACCCACAGAAAGAACCGAATCATGATAGTCCAGGCA 543

Query Match 88.7% Score 2065; DB 9; Length 3149;  
Best Local Similarity 99.2% Pred. No. 0; Mismatches 13; Indels 4; Gaps 3;  
Matches 2105; Conservative 0;

4 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 63  
211 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 270  
64 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAACACAGGCCAGGCCCAACCCAG 123  
271 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAACACAGGCCAGGCCCAACCCAG 330  
124 CCAGCCAAAGCAGCAGCAGCAACCCCGCCCGCCGAGAGACCTCCCAACCTTAACAGCCC 183  
331 CCAGCCAAAGCAGCAGCAGCAACCCCGCCCGCCGAGAGACCTCCCAACCTTAACAGCCC 390  
184 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 243  
391 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 450  
244 CACCAAGTTGCAATGGCTTTCCAGCAGCCTGTGATGCCCTCAAGCTGAACTCCTGAT 303  
451 CACCAAGTTGCAATGGCTTTCCAGCAGCCTGTGATGCCCTCAAGCTGAACTCCTGAT 510  
304 TACTTAAGATCAATTAAGAGCCTATGATGGGAACAATAAGAGAGGCTTGGAAAAC 363  
511 TACTTAAGATCAATTAAGAGCCTATGATGGGAACAATAAGAGAGGCTTGGAAAAC 570  
364 AACTATTACGAGATGCTCAGGATGATATCCAGACTTCAACACTATGTTTCAAAATGT 423  
571 AACTATTACGAGATGCTCAGGATGATATCCAGACTTCAACACTATGTTTCAAAATGT 630  
424 TACATCTACACAAAGCCTGAGAGATGATGTTTAATGGCAGAACTCTGAAAAAGCTC 483  
631 TACATCTACACAAAGCCTGAGAGATGATGTTTAATGGCAGAACTCTGAAAAAGCTC 690  
484 TTCTTGGCAAAAAATAATAGCTACCCACAGAAAGAACCGAATCATGATAGTCCAGGCA 543

691 TTCTTGGCAAAAAATAATAGCTACCCACAGAAAGAACCGAATCATGATAGTCCAGGCA 750  
544 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 603  
751 AAGAGGACAGCAGCAGCAGCAACCTTCACTGCTCAGAGTGTCTCAAGACATATGGAAA 810  
604 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 663  
811 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 870  
664 CCGTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 723  
871 CCGTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 930  
724 CCGTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 783  
931 CCGTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 990  
784 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 843  
991 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1050  
844 GGCACCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 903  
1051 GGCACCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1110  
904 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 963  
1111 CCAAAACACACTCAAGATGATGATGCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1170  
964 AAGCTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1023  
1171 AAGCTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1230  
1024 GACTCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1083  
1231 GACTCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1290  
1084 TGCAGGAGCATCTCAAGAGATGTTTGGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1143  
1291 TGCAGGAGCATCTCAAGAGATGTTTGGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1350  
1144 TCAAGAGCTGTGAGAGTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1203  
1351 TCAAGAGCTGTGAGAGTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1410  
1204 CCCATGAGATGAGCAGCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1263  
1411 CCCATGAGATGAGCAGCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1470  
1264 GAGTTTGTGCTGACGTCGATGATGTTCTTCAACTGCTATTAAGTCAAACTCCTCTGAC 1323  
1471 GAGTTTGTGCTGACGTCGATGATGTTCTTCAACTGCTATTAAGTCAAACTCCTCTGAC 1530  
1324 CATGAGAGTGTGAGAGTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1383  
1531 CATGAGAGTGTGAGAGTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1590  
1384 ATGCGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1443  
1591 ATGCGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1650  
1444 ACCAAGGTTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1503  
1651 ACCAAGGTTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1710  
1504 GACAGTTGAGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1563  
1711 GACAGTTGAGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1770  
1564 CAGCTCAAAAGCGGTGACAGAGAGCAGTGTGAGAGCCTCTCTCAGCAGCAGCAGCAGCAGCAG 1623  
1771 CAGCTCAAAAGCGGTGACAGAGAGCAGTGTGAGAGCCTCTCTCAGCAGCAGCAGCAGCAGCAG 1830





QY	544	AAAGGAGAGAGACGTGGGAGAGAAAGAAACAGGAGACAGAAACCTGGGCTTTCACGGTA	603
Db	751	AAAGGAGAGAGACCTGGGAGAGAAAGAAACAGGAGACAGAAACCTGGCTTTCACGGTA	810
QY	604	CCAAACACAACTCAAGCATGACTCCTCCGACAGACCAGACCCCTCAGCCGATCTCTCT	663
Db	811	CCAAACACAACTCAAGCATGACTCCTCCGACAGACCAGACCCCTCAGCCGATCTCTCTCT	870
QY	664	CCCTGTGACAGGACACGGCTCACCCCTTCCCTGGCTGTACCCCGGACCTCATGTCTCAGAC	723
Db	871	CCCTGTGAGGACACGGCTCTACCCCTTCCCTGGCTGTACCCCGGACCTCATGTCTCAGAC	930
QY	724	CCCTGTCAATGACAGTGGGTGCTCCCGACCACTGCACAGCCCGCGAGTGGCCCCCGAG	783
Db	931	CCCTGTCAATGACAGTGGGTGCTCCCGACCACTGCACAGCCCGCGAGTGGCCCCCGAG	990
QY	784	CCACAAACCCCAACCCGCTCCAGCTCCCCAGCCCGTACAGAGCCACCCACCCATCATGTGCG	843
Db	991	CCACAAACCCCAACCCGCTCCAGCTCCCCAGCCCGTACAGAGCCACCCACCCATCATGTGCG	1050
QY	844	GCCACCCCAACAGCCTGTGAAGACAAAGAGAGAGTGAAGAGGAAAGAGACACACACAC	903
Db	1051	GCCACCCCAACAGCCTGTGAAGACAAAGAGAGAGTGAAGAGGAAAGAGACACACACAC	1110
QY	904	CCGACACACAAATGACCCCAATTCAGAGGCCAACCTCGCTGCTCCCGGAGGCCAAAGCAC	963
Db	1111	CCGACACACAAATGACCCCAATTCAGAGGCCAACCTCGCTGCTCCCGGAGGCCAAAGCAC	1170
QY	964	AAAGTGGGGACAGCGCGCGGGAGACACCGCGCTGTGAACCTTCACAAAGAGACGCTGGCC	1023
Db	1171	AAAGTGGGGACAGCGCGGGAGACACCGCGCTGTGAACCTTCACAAAGAGAGCTGGCC	1230
QY	1024	GACTCTCAGACACACCCACGACACAGAGAAAGACAGCAAGGTCTCGAGACGCTCAATGTC	1083
Db	1231	GACTCTCAGAGACACCCACGACACAGAGAAAGAGAGCAAGGTCTCGAGACGCTCAATGTC	1290
QY	1084	TGCGAGGGGAAATCCCAAGAGAGATGTTGGCAAGAACACAGCGGCTACGGCTGGGCTTC	1143
Db	1291	TGCGAGGGGAAATCCCAAGAGAGATGTTGGCAAGAACACAGCGGCTACGGCTGGGCTTC	1350
QY	1144	TACAAGCCTGTGACGCTGAGAGGACACTGGGCTTACACAGCATCTGTGACATCTCAAGAC	1203
Db	1351	TACAAGCCTGTGAGCGTGGAGCGTGGAGGACACTGGGCTTACAGACTGTGTGACATCTCAAGAC	1410
QY	1204	CCCATGGACATGAGCACAATCAAGTCTAAACTGTGAGGCCCGGTGAGTACCGTGATGCTCAG	1263
Db	1411	CCCATGGACATGAGCACAATCAAGTCTAAACTGTGAGGCCCGGTGAGTACCGTGATGCTCAG	1470
QY	1264	GAGTGTGTCTGACGTCGCAATGATGTCTCCAACTGCTATAAATACAAACCTCTCTAC	1323
Db	1471	GAGTGTGTCTGACGTCGCAATGATGTCTCCAACTGCTATAAATACAAACCTCTCTCTAC	1530
QY	1324	CATGAGCTGTGTGCCATGATGCGCCGACAGCTTCAGAGATGTGTTGGAATGCGCTTGGCCAG	1383
Db	1531	CATGAGCTGTGTGCCATGATGCGCCGACAGCTTCAGAGATGTGTTGGAATGCGCTTGGCCAG	1590
QY	1384	ATGCGGAGAGAGCCTGTGAGAGCCAGTGTGTGGCCGTCTCCCGGGCAGTGGCCCCCTCC	1443
Db	1591	ATGCGGAGAGAGCCTGTGAGAGCCAGTGTGTGGCCGTCTCCCGGGCAGTGGCCCCCTCC	1650
QY	1444	ACCAAGTTTGTGGCCCCCGCTTATCCACAGCAGACAGAGAGAGGATAGTCTCTCGACAGT	1503
Db	1651	ACCAAGTTTGTGGCCCCCGCTTATCCACAGCAGACAGAGAGAGGATAGTCTCTCGACAGT	1710
QY	1504	GACAGTTTCGACTATGTACTCTGTGAGAGAGAGCGAGCCACAGCGCTGGGTGAGGCTCCAGAG	1563
Db	1711	GACAGTTTCGACTATGTACTCTGTGAGAGAGAGCGAGCCACAGCGCTGGGTGAGGCTCCAGAG	1770
QY	1564	CAGCTCAAAAGCGTGCACGACAGCTTGTGCAGGCTCTCTCAGCCGCCACAGCAACAAACA	1623
Db	1771	CAGCTCAAAAGCGGTGCACGACAGCTTGTGCAGGCTCTCTCAGCCGCCACAGCAACAAACA	1830
QY	1624	AAAGAAAAGAGAAAGACAGAAAGAAAGAAAGAAAGAAAGACACAAAGAGAAAGAGAA	1683

D	b	1831	AAAGAAAAAGGAAAAGCAACAGAGAAAAAAGAAAAAGCACAAGAAAAAGGAAGA	1890
O	y	1684	GTCGAAGAGAAATAAAAAAGCAAAGCCCAAGAACCTCCTTAAAAAGCAGACAGAAAAAT	1743
D	b	1891	GTCGAAAGAGAAATAAAAAAGCAAAGCCCAAGAACCTCCTTAAAAAGCAGAGAAAAAT	1950
O	y	1744	AATAGCACACACAGCATGTGTAGCAAGAAAGAGCGGCCCATGTAGAGACAGCCCCCT	1803
D	b	1951	AATAGCACACCAACAGCAAAATGTAGCAAGAAAGAGCGGCCCATGTAGAGACAGCCCCCT	2010
O	y	1804	CCACGATTAGTGTCGGAGGAAGAGCAAGTAGTCAGAACCTATGTCTTATGAGAGAGACGG	1863
D	b	2011	CCCACGTTAGTAGTCGGAGGAAGAGCAAGTAGTCAGAACCTATGTCTTATGAGAGAGACGG	2070
O	y	1864	CAGCTCAGCTTGGACATCAACAGAGTCCCCCGGCAGAGAGCTGGGCCCGCTGTGCACATC	1923
D	b	2071	CAGCTCAGCTTGGACATCAACAGAGTCCCCCGGCAGAGAGCTGGGCCCGCTGTGCACATC	2130
O	y	1934	ATCCAGTCAGCGGAGACCCTCCCTGTAAGATTCACACCCGACAGAGATTGAATGACTTTT	1983
D	b	2131	ATCCAGTCAGCGGAGACCCTCCCTGTAAGATTCACACCCGACAGAGATTGAATGACTTTT	2190
O	y	1984	GAGACCCGTAAGCCGCTCCACACTGGGTAGCTTGGAGCGCATGTACACTCTGTTCGG	2043
D	b	2191	GAGACCCGTAAGCCGCTCCACACTGGGTAGC-TGGAGCGCATGTACACTCTGTTCGG	2249
O	y	2044	GAAGAAAAAGAAACCTTCCACCTGAGAAAGTTGATGTATGTGCGGGTCTCCACANAT	2103
D	b	2250	GAAGAAAAAGAAAC-TCACACTGAGAAAGTTGATGTAT--GCCGCTCTCTCCACANAT	2306
O	y	2104	GAAGGNTTCTCGGTCTTCAAG	2125
D	b	2307	GAAGGCTTCTCGTCTCCAGAG	2328
RESULT_3				
LOCUS		AF461396	2727 bp mRNA linear ROD 13-MAY-2002	
DEFINITION			Mus musculus bromodomain-containing protein BRD4 short variant	
ACCESSION		AF461396	(Brd4) mRNA, complete cds.	
VERSION		AF461396.1	GI:18308126	
KEYWORDS				
SOURCE				
ORGANISM			Mus musculus.	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE			Wilson, S.T., Bullock, S.L., Lynch, D.E., Grigorjeva, E.F.,	
JOURNAL			Houzelstein, D., and Beddington, R.S.	
MEDLINE			Growth and early postimplantation defects in mice deficient for the	
PUBMED			bromodomain-containing protein Brd4	
AUTHORS			Mol. Cell. Biol. 22 (11), 3794-3802 (2002)	
JOURNAL				
TITLE				
FEATURES				
SOURCE				
gene				
CDS				
/product=			"Bromodomain-containing protein BRD4 short	





QY 781 CAGCCACACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 840  
 DB 1010 CAGCCACACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 1069  
 QY 841 GCGGCCACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 900  
 DB 1070 GCGGCCACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 1129  
 QY 901 ACCCCACACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 960  
 DB 1130 ACCCCACACCCCGCTCCAGCTCCAGCCCGTACAGAGCCACCCATCATC 1189  
 QY 961 ACACAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 1190 GCGGAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249  
 QY 1021 CCGGAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 DB 1250 CCGGAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1309  
 QY 1081 TGTGCGACGCGCATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 DB 1310 TGTGCGAGTGGCATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1369  
 QY 1141 TTCTACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 DB 1370 TTCTACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429  
 QY 1201 CACCCCATGAG 1260  
 DB 1430 CACCCCATGAG 1489  
 QY 1261 CAGAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 DB 1490 CAGAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549  
 QY 1321 GACCATAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 DB 1550 GACCATAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1609  
 QY 1381 AAGATGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1610 AAGATGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669  
 QY 1441 CCCACAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 DB 1670 CCCACAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729  
 QY 1501 AGTACAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 DB 1730 AGTACAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789  
 QY 1561 GAGCAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 DB 1790 GAGCAGAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849  
 QY 1621 CCAAG 1680  
 DB 1850 CCAAG 1909  
 QY 1681 GAAAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 DB 1910 GAAAGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969  
 QY 1741 AATTAATGAG 1800  
 DB 1970 AATTAATGAG 2029  
 QY 1801 CCTCCACAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 DB 2030 CCTCCACAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2089  
 QY 1861 CCGCAGCTGAGCTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920

DB 2090 CCGCAGCTGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2149  
 QY 1921 ATTCATGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 DB 2150 ATTCATGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2209  
 QY 1981 TTTGAGAGCTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
 DB 2210 TTTGAGAGCTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2268  
 QY 2041 GCGGAG 2100  
 DB 2269 GCGGAG 2325  
 QY 2101 NATGAAGGNTTCGGTCTTCAG 2125  
 DB 2326 GATGAAGGATTCATCTCTGAG 2350

RESULT 5  
 AF273217  
 LOCUS 5281 bp mRNA linear ROB 28-AUG-2000  
 DEFINITION Mus musculus cell proliferation related protein CAP (Cap) mRNA.  
 ACCESSION AF273217  
 VERSION AF273217.1 GI:9931485  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 5281)  
 Dev.A., Ellenberg,J., Farina,A., Coleman,A.E., Meryama,T.,  
 Sciorfino,S., Lipincott-Schwartz,U. and Ozato,K.  
 A bromodomain protein, MCPAP, associates with mitotic chromosomes  
 and affects G(2)-to-M transition  
 Mol. Cell. Biol. 20 (17), 6537-6549 (2000)  
 20396330  
 MEDLINE 10938129  
 PUBMED

REFERENCE  
 AUTHORS Dey,A., Bhattacharya,N., Horn,V., Loneragan,M., Cheng,K. and Ozato,K.  
 TITLE Submitted (30-MAY-2000) Lab. of Molecular Growth Regulation,  
 NICHD/NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA  
 JOURNAL  
 FEATURES  
 source  
 1..5281  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="17"  
 1..5281  
 /gene="Cap"  
 181..4383  
 /note="chromatin associated phosphoprotein; Bromodomain  
 and Et domains"  
 /product="cell proliferation related protein CAP"  
 /codon\_start=1  
 /protein\_id="AA02191.1"  
 /db\_xref="GI:9931486"  
 /translation="MSTESGPGTRLRNLPVMDGLFSONSTTQAOAPOPANASTN  
 PPPETSNPKPKRQTNOLYLRYVLTLMKQFAPFQPDVAKLNPVYKLIK  
 TPDMDGTLKRLNNYVWNAOECIDPENTFTQCYIYKRGDILVMAEALBKLFIQK  
 INELPPEEFIMIVQAKRGGRGRKELGAAPGVSTVPTNTQASTSPOTOTPOONPP  
 VQATDPFPAVPTDILAQPPVMTVPVPPQIQTPSPVPVPPPPPAVPPQVQSHPTI  
 ATPQPVKTKKGRKADITPTTIDIEPPLAPETKALGPRESSRPVPPK  
 DIVDSQHPGPKSSKISQKCCSGILKEMFKKNAVAAMPYKPVDEVALDHLQY  
 DLKHPMDSTIKSLSEYRDAQGVADRLMFSYKPNPDEVVAMARKLDY  
 FEMRPKMPDEPEPVYVSSPAVPPTKTVADPSSSDSSSDSSDSTDSSEER  
 AORLAELQDLKAVHQLAALSDPQONPKPKKKKKKKKKKKKKKKKKKKKKKK  
 KELPKKTKNNNSNSVSKKEPVPTKTPPTPYESEEDKCKPKSTEEKRLSLDIN  
 KLPGKLVGVYHIIQSREPSLKNSEDEIDETLKPSTLRLELYVTSCLKRRKP  
 QAEKVVIAGSRKMFSSSESTSESSSSDSEDETEMAKPKKKGHGRDQKHH



Qy	1921	ATCATTCAGGTACCGGGAGCCCTCCCTGAAAGAAATTCACACCCCGACAGACATTTGAAATGCAC	1980
Db	2089	ATTATTTTCAGTCAAGGGAAACCATCTATTTAAAACTCCAAACCCGATGAGATTTGAGATTTCAC	2148
Qy	1981	TTTTCAGACCTCTGAAGCCGCTCCACACTGCGTGAGCTTGGAGCGCTATGTCACCTCCTGTTT	2040
Db	2149	TTTTCGAGACCTCTGAAGCCATCTTCACACTTCACGAGAG--TTGAGCGGATATATGTCACCTCCTGTTT	2207
Qy	2041	GGCGAAGAAAAGAAACCTTCAGAGCTGAGAAAGTTGATGTGATNTGCCGGGTTTCCTCCAA	2100
Db	2208	GGCGAAGAAAAGAAAGAAC--TCAGAGCTGAGAAAGTTGACAGTATTT--GCTGTTCTTCCAA	2264
Qy	2101	NATGAAGGNTTCGCGTCTCAAG	2125
Db	2265	GATGAAGGATTTCTATCTCTGAG	2289

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION
BC008532	BC008532	Mus musculus, clone IMAGE:3489640, mRNA.	BC008532	BC008532.1
		2199 bp		GI:14250219
		linear		
		ROD 07-AUG-2002		

**SOURCE ORGANISM**

REFERENCE  
1 (bases 1 to 2199)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

FILE  
JOURNAL

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK	COMMENT
NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>	
Contact: MGC help desk	

Email: Cgaps@remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) Genome  
DNA Sequencing by: Sequencing Group at the Stanford Human  
Center, Stanford University School of Medicine, Stanford, CA 94305  
<http://www.shgc.stanford.edu>  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [modexall@stanford.edu](mailto:modexall@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 7 Row: b Column: 12  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarly but not identically to proteIn.

**FEATURES**  
**SOURCE**

```

7.0.553
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:3489640"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old. gross tissue."
/clone_id="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT      611 a      642 c      542 g      404 t
ORIGIN

```

Query Match	53.9%	Pred. 1256.2	DB 10	Length 2199
Best Local Similarity	88.6%	Pred. NO. 1.7e-26		
Matches 1373; Conservative	0	Mismatches 173;	Indels 3	Gaps 1;

OY	119	CCGAGCCAGGCAACGGCAGCCAGCAACCAACCCCGCCCGCCAGAGACCTCCAAACCCCTAAC	178
Db	647	CTGGCAGCCCTTGAGGATGTGGAAACCACTAGGGGGCTCCAGAGACCTCCAAACCCCTAAC	706
OY	179	AGCCCAAGAGGCAAGCCAAACCAACTGCATATACCTGCTCAGATGGTGTCTCAAGACATAT	238
Db	707	AGCCCAAGAGCAAGCAACCAACCACTGCATATATCTGCTCAGAGATGGTGTCTCAAGACATAT	766
OY	239	GGAACACACAGTTTGCATGGCCCTTTCAGCAGCCCTGTGGATGCCGTCAGACTGAACCTCC	298
Db	767	GGAACACACAGTTTGCATGGCCCTTTCAGCAGCCCTGTGGATGCCGTCAGACTGAACCTCC	826
OY	299	CTGATTAATATTAAGATCATTTAAACCCCTATATGGATATGGCAACATATAAAGACCCCTTG	358
Db	827	CTGATTAATATTAAGATTAATTTAAACACCCATGGATATGGCAACATATAAAGACCCCTTG	886
OY	359	AAACAACTATTAATCTGGAATGCTCAGAGATGTATCCAGACTTCAACACTATATTTATCA	418
Db	887	AAACAACTATTAATCTGGAATGCTCAGAGATGTATCCAGACTTCAACACTATATTTATCA	946
OY	419	ATTGTACATCTACAAACAGCCCTGGAGATGACATAGTCTTAAATGGCAGAGAGCTTGGAAA	478
Db	947	ATTGTACATCTATTAACAGCCCTGGAGATGACATAGTCTTAAATGGCAGAGAGCTTGGAGA	1006
OY	479	AGCTCTTCTTGCAAAAAATTAATGAGCTACCCACAGAAACCCGATCATGATATGCC	538
Db	1007	AGCTCTTCTTGCAAAAAATTAATGAATGCTCTACAGAAACCTGATCATGATATGCTC	1066
OY	539	AGGCAAAAGGAAGGAGCGTGGGAGGAAAGAAACAGGAGACAAACCTGGCGTTTCA	598
Db	1067	AGGCAAAAGGAAGGAGCGGAGGAGGAAAGAAACAGGAGACAAACCTGGGTATATCA	1128
OY	599	CGGTACCAAAACAACTCAAGCATTCGACTCTTCCGAGACCCGACCCCTTACGCCGA - -	656
Db	1127	CGGTACCAAAACAACTCAAGCATTCGACTCTTCCGAGACCCGACCCCTTACGCCGAGC	1186
OY	657	-TCTCTCTCTGTGACAGGCCACGCCCTTACCCCTTCTCTGCGCTCACCCGGACCTCATCG	715
Db	1187	CTCTCTCACCTGTGACAGGCCCAACCACTACCCCTTCTCTGCTGTCAACCCAGACCTCATG	1246
OY	716	TTCAGACACCCGTCATATACAGTGTGCTGCCGACCATCTGGAGAGGCCCGCCGACATGC	775
Db	1247	CCGAGCCTCTGTATATGACATATGTGCCCCCTCAGCACTTCAGATCTCTTACCGGTAC	1306
OY	776	CCCCCAGCACAACCCCCACCCGCTCCAGCTCCAGCCGCTACAGAGCCACCCACCCA	835
Db	1307	CCCCCAGCACAACCCCCACCCGCTCCAGCTCCAGCACTTCAGATCTCTTACCGGTAC	1366
OY	836	TCATTCGGGCGACCCCAAGCAGCTGTGAAGACAAAGAGGAGTGAAGAGAAACGACACA	895
Db	1367	TCATTTGGAGACACCCCCCAAGCAGCTGTGAAGACAAAGAGGAGTGAAGAGAAACGACATA	1428
OY	896	CCACACACCCACACACCATTTGACCCCATTCACAGGCCACCCCTGCGTCCCGGAGGCCA	955
Db	1427	CCACACACCCCTACACACCATTCGACCCCATTCATGAGCACCCCTACTGGCCCGACGACCCA	1486
OY	956	AGACACCAAGCTGTGGGCAAGCGGCGGAGAGCAGCGGCGCTTGAACCTCCAAAGAGG	1015
Db	1487	AGACACCGCAAGCTGTGGGCTCTCGGGGGAGAGCAGCAGACTTGAAGCCTCCAAAGAGG	1546
OY	1016	ACGTGCCGCACTCTCAGCAGCACCCACGACACAGAGAAAGCAGCAAGGCTTTCGAGCAGC	1075
Db	1547	ATGTACCGGAGCTACAGCAGCACCCAGAGGCGAGAGAAAGAGCAGCAAGATCTGTGAGCAGC	1606
OY	1076	TCAGTGTCTGAGGGGATCTCCTCAAGAGATGTTTCCAAAGAACGCCGCTTACGCT	1135
Db	1607	TAAAGTCTCTCAGGGGATCTCTCAAGAGATGTTTCCAAAGAACATCTGTGCTTACGCT	1666
OY	1136	GGCCCTTCTCAAGAGCCTGTGGAGCTGGAGAGCACTGGGCTTACAGACTACTGTGACATCA	1195
Db	1667	GGCCCTTCTCAAGAGCCTGTGGAGATGTGGAGGCACTGGGTCTGACACACTACTGTGACATCA	1726
OY	1196	TCAAGCACCCCATGAGCATGAGCACAATCAAGTCTAACTGAGAGCCGCTGAGTACCGCTG	1255



Db	1727	TCAAACATCCCATGAGCATGAGCACAAATCAAGTCTAAACTAGAGTCCGAGAGTACAGAG	1786
QY	1256	ATGCTCAGGAGTTTGGTGTGACGTCCGATTGATGTTCTCCAACTGCTATTAAGTACAAACC	1315
Db	1787	ATGCCAGAGAAATTTGGTGTGTGATGTCCGATTTGATGTTCTCCAACTGCTCAAGTACAAAC	1846
QY	1316	CTCCGAGACATGAGTGGTGGTGGCCATGGGCCCCGAAGCTCCAGATGTGTTGCAAAATGCGT	1375
Db	1847	CCCCGACCATTAAGTATGGTATGACCATAGCGCTCGAAAACTCCAGAGATGTTTGGAAATGCCGT	1906
QY	1376	TTGCCAAGATGGCGGACGAGCCTTGAGAGACCAAGTGGTGCCGCTGTCCTCCCGGAGTGC	1435
Db	1907	TTGCCAAGATGCTCTATGAGCCTGGAAGACCAAGTTGTTACAATGTCTCTCTCGACGTGC	1966
QY	1436	CCCCCTCCACCAAGGTTTGCGCCCGCCCTCATCCAGCGACAGCAGACGATAGCTCCT	1495
Db	1967	CACCCCTTACAAAGGTGGTGAAGCCCAACCTCATCTATGATGACGACGACGACAGATGTTT	2026
QY	1496	CGGACATGAGACATTTGCACTGATGACTCTGAGAGAGAGCCGACCCGAGCGGCTGGCTGACC	1555
Db	2027	CCGAGAGCGACAGTTTCCACTGACACTCTGAGAGAGAGGAGCCGACCGGCTGGCTGAAC	2086
QY	1556	TCCAGAGACAGCTCAAAGCCGTGACAGACAGACAGCTTGCAGCCCTTCTCTAGCCCCACAGA	1615
Db	2087	TCCAGAGACAGCTCAAAGCCGTGACAGACAGACAGCTTGCAGCCCTTCTACAGCCCCACAGA	2146
QY	1616	ACAAACCAAAGAAAGGAGAAAGACAAGACAAGAAAGAAAAAGAAAA 1664	
Db	2147	ACAAACCAAAGAAAGGAGAAAGACAAGACAAGAAAGAAAAAGAAAA 2195	

RESULT	7
HUMORFX	
LOCUS	3028 bp mRNA linear PRI 06-OCT-2001
DEFINITION	Human mRNA for KIAA0043 gene, complete cds.
ACCESSION	D26362
VERSION	D26362.1 GI:452518
KEYWORDS	KIAA0043.
SOURCE	Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
ORGANISM	Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1
AUTHORS	Nomura,N., Nagase,T., Miyajima,N., Sazuka,T., Tanaka,A., Sato,S., Seki,N., Kawarabayashi,Y., Ishikawa,K. and Tabataba.S. Prediction of the coding sequences of unidentified human genes. II The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1
TITLE	DNA Res. 1 (5), 223-229 (1994)
JOURNAL	96051398
MEDLINE	2 (bases 1 to 3028)
REFERENCE	Ohara,O., Nagase,T., Kikuno,R. and Nomura,N. Direct Submission
AUTHORS	Submitted (27-DEC-1993) Osamu Ohara, Kazusa DNA Research Institute;
TITLE	1332-3, Yana, Kisanazu, Chiba 297-0812, Japan
JOURNAL	(E-mail:cdbnainfo@kazusa.or.jp, Tel:+81-438-52-3913)

FEATURES	
source	location/Qualifiers
	1..3028
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/sex="male"
	/cell_line="KG-1"
	/cell_type="myeloblast"
gene	1..3028
	/gene="KIAA0043"
5'UTR	1..139
	/gene="KIAA0043"
CDS	140..2320
	/gene="KIAA0043"
	/note="similar to Human homolog of Drosophila female sterile homeotic mRNA ( HMF5HG)"
	/codon_start=1

[illegible]



```

Db      830 CCTCTACGCGCGCTGTGTGAAGAAAAGGCGGAAAGCGAAGACACAACCACT 889
Qy      904 CCCACCAACATGAGACCCCATTCAC-----GAGCCACCCCTGCTGCCCCGGAGACC 954
Db      890 CCCAGAGAGTGGGCGATCAGCAGCGAGAGTGTGCGCCCGCGCTTGTACAGACC 949
Qy      955 AAGACCAACAGCTGGGCGACGCGGAGAGCA--GCGGCGCTGTGAAACCTCCAAAG 1011
Db      950 AAGCAGGCGCAAGGTGGCGCGCGGAGAGTGTGCGCCCGCCATCAACCTCCCAAG 1009
Qy      1012 AAGCAGGCGCGGAGCTGTCAACAGACCCAGACCAAGAAAGAGCAGCAAGTCTGGAG 1071
Db      1010 AAGCAGCTGGAGAGCGCGGAGCGCCAGACCGAGGCAAGAGAGGCACTGTGGAG 1069
Qy      1072 GAGCTCAAGTGTCTCAGCGGAGCTCTCAAGAGAGTGTGGCAAGAACACGCGGCTAC 1131
Db      1070 CAGCTGGGCTACTGTGAGAGAGATCTCTCAGGAGAGTGTCTCAAGAAAGACAGCGGCTAC 1129
Qy      1132 GCGTGGCGCTTCTACAAAGCTGTGAGCGTGAAGGCACTGGGCTACAGCAGTGTGAC 1191
Db      1130 GCGTGGCGCTTCTACAAAGCTGTGAGCGTGAAGGCGCTGTGAGCAGTGTACACAGAC 1189
Qy      1192 ATCATCAAGCAACCCCATGAGATGAGACATCAAGTCTAACTGGAGGCGCGTGAATAC 1251
Db      1190 ATCATCAAGCAACCCCATGAGATGAGACATCAAGTCTAAAGAGATGAGGCGGAGATAC 1249
Qy      1252 CGTATGCTCTCAGAGATTGTGTGAGCTGAGCTGAGTGTGATGTTTCAACTGCTATTAAGTAC 1311
Db      1250 CCACAGCAGAGGCGCTTGTGCTGATGTGTGCGGTGATGTCTGAAATGTCTACAAATAC 1309
Qy      1312 AACCTCTCTGACATGAGAGTGTGCGCATGAGCGGCAAGCTCAAGATGTGTGAAATG 1371
Db      1310 AATCCCCAGACACGAGGTGTGTGCGCATGAGCGGCAAGCTCAAGATGTGTGAGATG 1369
Qy      1372 CGTGTGCAAGATGCGCGGAGAGCTGAGAGGCAAGTGTGCGGTCTCTCCCGGCA 1431
Db      1370 AGGTTGTGCAAGATGCGCGGAGAGCTGAGAGGCAAGTGTGCGGTCTCTCCCGGCA 1417
Qy      1432 GTGCGCGCTCTCCACCAAGATTGTGTGCGCGCGCTCTATCAGACGACGACGAGCATAGC 1491
Db      1418 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1477
Qy      1492 TCTCTGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1551
Db      1478 TCTCTGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1531
Qy      1552 GAGCTTCAGAGAGAGCTCAAGAGCGCTGACAGAGAGCTGTGAGCGCTCTCTCAGCGCCAG 1611
Db      1532 GAGCTTCAGAGAGAGCTCAAGAGCGCTGACAGAGAGCTGTGAGCGCTCTCTCAGCGCCAG 1591
Qy      1612 CAGAACCAACCA-----AGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
Db      1592 GTAAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
Qy      1660 GAAAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719
Db      1652 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
Qy      1720 CCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1779
Db      1712 CCGCTCTCCAGAGAGAGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
Qy      1780 GCGCCCATGAG-----AGCAAGCGCGCTCTCCAGCTATGAGTGTGAG 1821
Db      1772 AGCGCGGAGAGAGAGTGAAGAAAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
Qy      1822 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881
Db      1832 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
Qy      1882 AACAAAGTCCCGGCGAGAGAGCTGGGCGGTGTGACATCACTCAGTCCGAGAGAGCC 1941

```

```

Db      1892 AACCGGCTGCCCGGAGAGAGCTGGGCGCGGTAGTGTGACATCACTCACTCGGAGACC 1951
Qy      1942 TCCCTGAAGAAATTCACACCCGAGAGATGAATGCACTTTGAGACCTGAGCGCTCC 2001
Db      1952 TCGCTCAGAGAGACTCCACCCGAGAGAGATGAATGCACTTTGAGAGCTGTGAACCCACC 2011
Qy      2002 AACTGCGTGAAGCTTGAAGCGCTATGTCACCTCTGTTTGGGAGAGAGAAAGCACTTC 2061
Db      2012 ACTTGTGCGGGAAC-TGAGAGATATATGTCAAGCTGTTTACAGAAAAGCAAGAGAAACC 2070
Qy      2062 AAGCTGAGAAAG 2073
Db      2071 GTTTCAGCAAG 2082

```

```

RESULT 8
LOCUS   BC031536
DEFINITION Mus musculus, bromodomain-containing 3, clone MGC:28328
ACCESSION BC031536
VERSION   BC031536.1 GI:21594630
KEYWORDS MGC.
SOURCE   house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3033)
AUTHORS Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK   NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgabs-rtmail.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU)
          Sequencing Center
          Center code: BCM-HGSC
          Web site: http://www.hgsc.bcm.tmc.edu/cdna/
          Contact: amg@bcm.tmc.edu
          Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
          Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
          Richards, S., Gibbs, R.A.
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/ILNU at: http://image.llnl.gov
          Series: IRAC Plate: 36 Row: m Column: 24
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 12963646.
          Location/Qualifiers
            1. 3033
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /map="C2BCH.11"
              /clone="MGC:28328 IMAGE:4015879"
              /tissue_type="Mammary tumor metastatized to lung.
              MMTV-LTR/mtl model. Expression driven by an MMTV-LTR
              enhancer."
              /clone_1lb="NCI CGAP Lu30"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6"
              167..2347
                /codon_start=1
                /product="bromodomain-containing 3"
                /protein_id="AAH31536.1"
                /db_xref="GI:21594631"
                /db_xref="LocusID:67382"
                /translation="MSTTAAPTGIPAVPGVPVNPPEVSNPSKPKRTINQLOVMNV

```















Db 1332 TCCGGAGCCTGTGGATGCTCAAACTGGGTCTACCGGTTATTCACAAAATTTATTAAC 1391  
 QY 323 CGCCTATGATGGAACAAATTAAGAGCGCTTGGAAAAACAATTTACTGGAATCTC 382  
 Db 1392 AGCCTATGGAATGGGTACTATTAAAGAGAGCTTGAACAAATTTATTTATGGGCTCT 1451  
 QY 383 AGGAATGTATCCAGAGCTTCAACATATGTTTCAAAATTTTTCATCTACACAGCCTG 442  
 Db 1452 CAAGAGTATGCAAGATTTTAAATACATGTTCCAACTGTTCAATTTTACAAAGCCCA 1511  
 QY 443 GAGATGACATAGTCTTAATGGCAGAAAGCTGTGAAAAAGCTCTTGGCAAAAATAAATG 502  
 Db 1512 CTGATATATTTGCTTAATGGCACAACGCTGAAAAAGATTTCTTACAGAAAGTTGCAT 1571  
 QY 503 AGCTACCCACAGAAAGAACGATCATGATATGTCAGGCAAGAAAGAGAGAGAGCTGGGA 562  
 Db 1572 CATGCCACAGAAAGAAAGAGAGCTGGTAGTACCATCCCTAAGAACAGCCACAAAGAG 1631  
 QY 563 GGAAGAAACAGGAGACAGCAAAACCTGGCGTTTCCAGGGTACCAAAACAACTCAAGCAT 622  
 Db 1632 GGGCCAAAGTTGGCAGGCTCCAGGCGAGTGTACAGTGCCCATCAGGTGCTGCGCTCT 1691  
 QY 623 CGACTCTCCGAGACCCGACCCCTCAGCCGCAATCTCTCTGTCAGAGGCGACGCTC 682  
 Db 1692 -----CTTCTGTGTACACACAGCCCTGTATATCTCTCAGCTGAGATACCTACACATG 1745  
 QY 683 ACCCTTCCCTGCGCTACCCCGAGACTCATGCTCCAGACCCCTGTATGATGAGTGGTGC 742  
 Db 1746 TCCCTCAACATVCCCAACCATCATGTCATTTCC-----T 1778  
 QY 743 CTCGCCAGCATGCGAGAGCGCCCGCCGAGTGGCCCCCAAGCCACAAACCCACCGCTC 802  
 Db 1779 CTCACATTTCTCAAGTCTTCTGACCTGTGAGACCCCGCTCTGCTGTATACGAGCTC 1838  
 QY 803 CAGCTCCAGCGCCGTACAGAGCCACCCATCATGCGGCGCCACCCACAGCTGTGA 862  
 Db 1839 CTCACAGCCAGCCCTTGGCAAGAAAAAGGCT----- 1872  
 QY 863 AGACAAAGAGGAGTGAAGAGAAAGACAGACACCAACCCACCCACACCATTTAGACCCA 922  
 Db 1873 ----AAAGGGAAGAGAGTACTACACCCCTACACCTACAGCATTCTGGCTCTGTT 1928  
 QY 923 TTCAGAGCACCCTGCTGCGCCCGGAGCCCAAGACACCAAGCTGGGCG----ACGCGC 979  
 Db 1929 CTCACACTAGCCCTCTGGGAGTCTTGAGCCTTAAGGACACAGCGCTTCCCTATCCGTA 1988  
 QY 980 GGGAGAGCAGCGCGCTGTGAACCTCTCAAAAGAGAGCGTCCGAGCTCTCAGCAGCAC 1039  
 Db 1989 GAGAGAGTGGTGGCCCATCAAGCCCCACGCAAGAACTTGGCTGACTCTCAGCAACAC 2048  
 QY 1040 CAGCACCAGAGAGAGCAGCAGAGTCTCGAGCAGCTCAAGTCTGACGCGCATCTCA 1099  
 Db 2049 ACCAGAGCTCTAAGAAAGAAAGCTTTCAGAACAGTTAAACATTCATGCGATTGTA 2108  
 QY 1100 AGGAGATGTTTGGCAAGAGCAGCGCGCTAGCGCTGGCCCTTCTCAAGACCTGTGGAG 1159  
 Db 2109 AGGATTTACTCTTAAGAGCATGCTGCTATGCTGGCTTCTTAAACCAAGTGGATG 2168  
 QY 1160 TGGAGCAGCTGGGCTTACAGCTACTGTGACTCATCAAGCAGCCCATGAGATGAGA 1219  
 Db 2169 CTTTGCACTTGGCTGCTGACTGACTACATCATTTAAGCAGCCCATGAGACTTACGA 2228  
 QY 1220 CAATCAAGTCTAAACTGAGGCGCGTGAATACCGTGAATGCTGAGAGTGTGCTGAG 1279  
 Db 2229 CTGTCAAGGGAAGATGAGAACCGTGAATACCGGATGACAGAGGATTTGCTGAG 2288  
 QY 1280 TCGGATGATGTTCTCAACTGCTATTAATACAAACCTCTGACCTGAGTGGTGGCCA 1339  
 Db 2289 TACGGTGTATGTTCTCAACTGCTATTAATACAAACCTCTGAGTGTGTTGCA 2348  
 QY 1340 TGGCCCGAAGCTCAGATGTTGTAATGCGCTTTCAGAAAGATGCGGAGAGCGACCTG 1399

Db 2349 TGGCAGAAAGTACAGGATGTATTTAGTTCGTTATGCCAAGATCCAGATGAACAC 2408  
 QY 1400 AGGAGCAGTGTGTGGCGCTGTCTCCCGGCACTGGCCCTCC---ACCAAGTTGTG 1456  
 Db 2409 TAGAACAGGCGCTTTACAGAGTCTTACTAGTCCATGGCCCTGGCTTGGCAATTCGTT 2468  
 QY 1457 CCCCCCTCATCCAGCAGCAGCAGCAGGATAGCTCTC----- 1496  
 Db 2469 CAGAGTCTTCAGTGAAGAAAGTACAGTGAAGAGCTCTCTGAGAAAGAGAGAGAA 2528  
 QY 1497 -----GGACAGTGAAGTTCGATGATGACTGTGAGAGAGAGAGAGAGAGAG 1540  
 Db 2529 ATGAGAGAGCAG 2588  
 QY 1541 AGCGCTGTGTGATCTTCAAGAGCAGCTCAAAACCTGTGACAGAGAGAGAGAGAGAGAG 1600  
 Db 2589 ATGCTTGAAG 2648  
 QY 1601 CTGAGCCCGCAG 1659  
 Db 2649 CCGAGGCTCCATATATCCAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2708  
 QY 1660 GAAAGCAGAAAG 1705  
 Db 2709 GGAAGGAG 2768  
 QY 1706 AAGCCAAAG 1749  
 Db 2769 CACCCCGCCACCTCAACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2828  
 QY 1750 -----AGCAAG 1771  
 Db 2829 CTGCTTAAAGCCCTTCTGCTTGTGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 2888  
 QY 1772 AGGAGCAG 1831  
 Db 2889 AGGCAACAAAG 2948  
 QY 1832 AGTCAAGCCTATGCTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891  
 Db 2949 AGAGCAGGCGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3008  
 QY 1892 CCGGCGAG 1951  
 Db 3009 CTGGGAG 3068  
 QY 1952 ATTCCAAAGCAG 2011  
 Db 3069 ATTCAAAAGCAG 3128  
 QY 2012 AGCTTGAAGCCTATGCTCACTCTCTGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2050  
 Db 3129 AGCTT-GAGCGGATGCT 3166  
 RESULT 14  
 HUMFSG 4053 bp mRNA linear PRI 22-JUN-1995  
 LOCUS Human homolog of Drosophila female sterile homeotic mRNA, complete cds.  
 DEFINITION  
 ACCESSION M80613  
 VERSION M80613.1 GI:182768  
 KEYWORDS MHC class II antigen; female sterile homeotic gene.  
 SOURCE Homo sapiens CDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 4053)  
 AUTHORS Beck,S., Hanson,J.M., Kelly,A., Paplin,D.J.C. and Trowsdale,J.  
 TITLE A homologue of the Drosophila female sterile homeotic (Fsh) gene in the class II region of the human MHC  
 JOURNAL DNA Seq. (1991) In press  
 REFERENCE 2 (sites)





Db	2709	GGAAAGCGAAGAGATCGAGGCGCAGGCTGGGGCCGAGCAAGATGACAAAGGGCGCTAAGG	2768
Qy	1706	AAGCCAAAGAACCTCTCTCTTAAAAAGACAGAAAAATATATAGC-----	1749
Db	2769	CACCCGCCACCTCAACCTAAGAGTCCAAAGAAAGCAAGTGGAGTGGGCTGGCAGTG	2828
Qy	1750	-----AGCAACGACATATGACCAAGA	1771
Db	2829	CTGCTTTAGGCCCTTCTGGCTTTGGACCTTCTGGAGAGAAATGGCACCAAGCTCCCAAAA	2888
Qy	1772	AGGAGCCAGCGCCCATGAAAGACAGCCGCCCTCCACGATAGAGTGGAGAAAGAGACA	1831
Db	2889	AGGCCACAAAGACAGCCACCCTGCCCTGCGCTACAGGTTATGATTTCAGAGGAGGAGGAAG	2948
Qy	1832	AGTGCAGACCTATGTCTCATGAGAGAAAGCGGCACCTAGCTTGGACATCAACAAGCTCC	1891
Db	2949	AGAGAGAGCCCTCTGATGATTACGATGAGAAAGCGGCACCTAGCCTGGACATCAACAATTTAC	3008
Qy	1892	CCGGGAGAGAGCTGGGGCGCGGTGGTGCACATATCATGACAGTACAGGAGGAGCCCTCCGAAAGA	1951
Db	3009	CTGGGGAGAGAGCTGGGGCCGAGTGTGTGCATATATATCAAGCCAGGGAGCCCTCTTTACGTG	3068
Qy	1952	ATTCCACCCCGCAGACAGATTGAAATGCACTTTGAGAACCCCTGAAAGCCGTCACACTGCGTG	2011
Db	3069	ATTCAACCCCAAGAGAGATTGATGATTGTTGAAACACTCAAGCCATCAACACTTAAG	3128
Qy	2012	AGCTTGAAGCCGTATGTCACTCTCTGTTTGGGGAAGAA	2050
Db	3129	AGCTT-GAGCGCTATGTCTCTTCTGCTCACTAGTAA	3166
RESULT 15			
LOCUS	AX409010	4664 bp	DNA linear PAT 14-JUN-2002
DEFINITION	Sequence 1657 from Patent WO0229103.		
ACCESSION	AX409010		
VERSION	AX409010.1	GI:21441715	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 Alvaras, C., Horne, D., Pears-da-Silva, S. and Vockley, J.G.		
TITLE	Gene expression profiles in liver cancer		
JOURNAL	Patent: WO 0229103-A 1657 11-APR-2002;		
FEATURES	GENE LOGIC INC (US)		
source	Location/Qualifiers		
	1..4664		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/note="EMBL/GenBank Accession No. D42040"		
BASE COUNT	1084 a 1336 c 1293 g 951 t		
ORIGIN			
Query Match	19.8%;	Score 460.2;	DB 6; Length 4664;
Best Local Similarity	57.0%;	Pred. No. 4.6e-90;	
Matches 1151; Conservative	0;	Mismatches 693;	Indels 175; Gaps 10;
Qy	143	CCAAACCCCGCGCCGACAGACCTCCAAACCTTACACAGCCCAAGAGGAGCAACCAAC	202
Db	1877	CCAAACCCCAACACCCCGGAGAGGTGTCCAAATCCCAAAAAGCCAGAGCAAGTTACCAACACAGC	1936
Qy	203	TGCATTAACCTGCTCAGAGTGTGTGCTCAAGACACTATGGAACACACAGTTTGGATGGCCTT	262
Db	1937	TGCATTAACCTACACAGATGATGATGAAGGCTGTGTGGAACATCATGTTCCGATGGCCAT	1996
Qy	263	TTCACAGCAGCTGTGTGATGCCGTCAAGCTGAACTCTCCGTGATTTACTATAAGATCATTAATAA	322
Db	1997	TCCCGCAGCCCTGTGTGATGCTGTCAAAAGGGGCTACCGGATTAATCAAAAATTTATAAAC	2056
Qy	323	CGCCATATGATATGGGAACATAAAGAAAGCGTTTGGAAAAACAATATTCAGTAATGCTC	382

D	2057	AGCTATGACAGTGGGATCACTATTAAAGAGGAGACTTGAATAACATTAATTATTGGGCTGCTT	2116
Q	383	AGGAATGTATCCAGGACTTCAACACTATGTTTACAAATTTGTTACATCTACAAACGCTG	442
D	2117	CAGATGTATGCAAGATTTTAAATACATGCTTCCACCAACTGTTACATTTTACACACGAAGCCCA	2178
Q	443	GAGATGACATAGTCTTAAATGGCAGAAAGCTCTGGAAAAGCTCTTCTTGCAAAATAAATG	502
D	2177	CTGATGATATTGTCTCTAATGGCAACAAACGCTGGAAAAGATTTCTTACAGAAAGGTTTGAT	2236
Q	503	AGCTAACCCAGCAAGAAACGAGATCATATGATTCAGAGCAAAAGAAAGAAAGAGACTGGGA	562
D	2237	CAATGCCACAAAGAAAGAACAGAGCTGGTAGTAGACATTCCTTAAGAACAGCCACAAAGAG	2296
Q	563	GGAAGAAGACAGGAGCAGCAAAACCTGGCCCTTTCCACGGTACCAACACACACTCAAGAT	622
D	2297	GGGGCAAGTTGGCAGCGCTCCAGGGCAGTGTACAGTCCCATCAGAGGCTGCGCTCT	2356
Q	623	CGATCTCCCGCAGACCCCAACCCCTCAGCGGAATCTCTCTCTGTGCAAGGCAAGCTC	682
D	2357	-----CTTCTGTGTCAACACAGACGCGCTGTATATCTCTCACCCTGAGATACCTACACTG	2410
Q	683	ACCCCTTCCCTGCGCCTACCCCGGAGCTCATGCTTCAGACCCCTGTGATCATGACAGAGTGTC	742
D	2411	TTCTCAACATTTCCCAACCCATCACTGATTTCC-----T	2443
Q	743	CTCCCAAGCCACTGACAGACGCCCCCGCAGTGGCCCCCGACAGCAACCCCAACCCGCTC	802
D	2444	CTCCACTTTCACAGTCTCTTCACTCTGCGAGAACCCCGCTCTGTGTACTGTGACAGCTC	2503
Q	803	CAGTCTCCCAAGCCCGTACAGAGCCACCCCAACCATCATGCGGGCCACCCCAACGCTGTGA	862
D	2504	CTCCAGCCCAAGCCCTTGGCCAAAGAAAAAGGGCT-----	2537
Q	863	AGACAAAGAGGAGTGTAAGAGAAAGAAAGACACACACCCCAACACATTTGACCCCA	922
D	2538	-----AAGCGGAAAGACGATACATCAACCCCTACACCTTACAGCATTTGGGCTCTGGTT	2596
Q	923	TTTCAGAGCAACCTGTGCTGCCCCCGGAGGCCAAGACACCAAGCTGGGGC---AGCGGC	979
D	2594	CTCCAGTACGCCCTTCCGGGAGTCTTGAAGCGTAAAGGACACAGCGCTTCCCTTATGGCTA	2653
Q	980	GGGAGACACACCCGCTGTAAACCTTCCAAAGAAAGAGCTGGCCGACTCTACAGACAC	1038
D	2654	GAGAGAGTGTGTGCCCCCATCAACCCCCCAAGCAAAACTGTGCTACCTCAGCAACAC	2713
Q	1040	CAGACACAGAAAGAGCAGCAGCAAGTCTCGAGAGCTCAAGTGTGCGACGCAATCTCA	1099
D	2714	ACCAAGACTCTAAGAAAGAAAGAAAGCTTTCAGAACAGTTAAACATTGCAATGGCATTTTGA	2773
Q	1100	AGGAGATGTTTGGCCAAAGAACAGCGCGCTTACGCCCTGTGACCTTCTTCAACGCTGTGACG	1159
D	2774	AGGAGTTACTCTCTAAGAAACATGTCGCCATATGCTTGGCTTTCTAATAACCAAGTGAATG	2833
Q	1160	TGGAGGACATGGGCTTACACGATTACTGTGACATATCAAGACACCCATGACATGACGA	1219
D	2834	CTTCTGCACTTGGCCTGCAAGACTTACATGACATCATCTTAAAGCACCCCATGACCTACACA	2893
Q	1220	CAATCAAGTCTAAACGCGAGGAGCGCCGAGAGACCGTATGCTCAGAGGTTTGGTGTAGC	1279
D	2894	CTGTCAACGCGAAAGATGAGAACCGTGTATTTACGGGATCACAAGAGTTTGTGTGTATG	2953
Q	1280	TCCGATTGATGTTTCTCCAACTGCTATATAGTACAAACCTCTGACCATAGAGTGGGCGCA	1339
D	3014	TGGCACAAGAAAGTACAGAGATGTATTGAGTTCGGTATGCCAAAGATCCAGATGAACAC	3073
Q	1340	TGGCCGCAAGCTCCAGAGATGTTTGAATGCGCTTGGCCAAAGATGCCGAGCAAGCCTG	1399
D	3074	TAGAAACAGGGGCTTTACCAAGTCTTACATGCGCAATGCGCCCTGTGGCGCAAAATGCTTT	3133



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:50:44 ; Search time 480 Seconds  
(without alignments)  
10926.884 Million cell updates/sec

Title: US-09-700-590A-101

Perfect score: 2329  
Sequence: 1 gcagggatcactagcatgctc.....ngccttaaaaggtlnccca 2329

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

N\_Geneseq\_101002:.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2316	99.4	2329	21	AA56719 Human transmembran
2	559.4	24.0	667	22	AA503029 Human diagnostic a
3	460.2	19.8	4664	24	ABR83660 Human cDNA differe
4	460.2	19.8	4664	24	ABN95159 Gene #1657 used to
5	371	15.9	1884	22	AAH18591 Human cDNA sequenc
6	294.2	12.6	3104	19	AAV68343 Transcriptional re
7	294.2	12.6	3106	20	AAK40172 WO9904265 Seq ID N
8	227.8	9.8	1862	23	ABY25740 Human prostate exp
9	224	9.6	424	23	AA526606 Human cDNA encodin

10	224	9.6	746	22	AA526193	Human cDNA encodin
11	208.2	8.9	1741	24	ABR84162	Human cDNA differe
12	207	8.9	352	22	AAF66021	Novel human polynu
13	203.8	8.8	6776	23	ABL04337	Drosophila melanog
14	168.8	7.2	209	20	AAV87340	EST clone BW2. Ho
15	159.4	6.8	560	24	ABO60244	Human colon cancer
16	159.2	6.8	450	21	AAZ80432	Human colon cancer
17	126.6	5.4	559	21	AAA43730	Mouse secreted exp
18	120.8	5.2	642	20	AAK29128	Polynucleotide RTP
19	92.8	4.0	597	23	ABV57572	Human prostate exp
20	85.2	3.7	807	23	ABL19177	Drosophila melanog
21	85.2	3.7	2807	23	ABL19176	Drosophila melanog
22	85.2	3.7	5017	23	ABL15790	Drosophila melanog
23	84.4	3.6	372	21	AAK59866	Eucalyptus grandis
24	79.2	3.4	140	19	AAV68386	Human BAZ gene pri
25	76	3.3	10629	22	AAH28279	- Nucleotide sequenc
26	75	3.2	2175	23	ABL12585	Drosophila melanog
27	75	3.2	4229	23	ABL12584	Drosophila melanog
28	75	3.2	10598	23	ABL27241	Drosophila melanog
29	75	3.2	15667	23	ABL27240	Drosophila melanog
30	71	3.0	1542	23	ABL23049	Drosophila melanog
31	71	3.0	3542	23	ABL23048	Drosophila melanog
32	69.8	3.0	1445	21	AAK47126	Arabidopsis thalia
33	69.6	3.0	2247	21	AAK6949	Arabidopsis thalia
34	68.6	2.9	2442	21	AAK50976	Arabidopsis thalia
35	68.4	2.9	306	21	AAK56643	Eucalyptus grandis
36	68.4	2.9	348	21	AAK55981	Eucalyptus grandis
37	68.2	2.9	693	23	AAK74240	DNA encoding novel
38	68.2	2.9	693	23	AAK90715	DNA encoding novel
39	68	2.9	1686	16	AAK87587	DNA encoding leuco
40	67.8	2.9	9443	23	ABL16121	Drosophila melanog
41	67	2.9	819	23	AAK68928	DNA encoding novel
42	66.8	2.9	441	23	AAK54543	DNA encoding novel
43	66.6	2.9	2443	24	ABO54948	Human ovarian anti
44	66.6	2.9	2444	21	AAK77905	Human cancer assoc
45	66	2.8	642	23	AAK71152	DNA encoding novel

## ALIGNMENTS

RESULT 1	
ID	AA56719 standard; cDNA; 2329 BP.
XX	AA56719;
XX	23-MAR-2000 (first entry)
DT	Human transmembrane protein HTMPN-22 encoding cDNA.
DE	
XX	Human: transmembrane protein; HTMPN; diagnosis; immunospecific;
KW	antiproliferative; neutroprotective; immune disorder;
KW	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	gastrointestinal disorder; developmental disorder;
KW	cell proliferative disorder; ss.
OS	Homo sapiens.
XX	
XX	WO9961471-A2.
FN	
XX	02-DEC-1999.
PD	
XX	
PF	28-MAY-1999; 99WO-US11904.
XX	
XX	29-MAY-1998; 98US-0087260.
PR	02-JUL-1998; 98US-0091674.
PR	02-OCT-1998; 98US-0102954.
PR	24-NOV-1998; 98US-0109869.
XX	(INCYT-) INCYTE PHARM INC.
PA	
XX	Tang YT, Lal P, Hillman JL, Yue H, Guegler KO, Corley NC;
PI	

PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;  
 PI Au-Young J;  
 XX WPI: 2000-072605/06.  
 DR P-PSDB: AAY57898.  
 XX  
 PT Proteins, polynucleotides, vectors, host cells and antibodies used to  
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,  
 PT neurological, gastrointestinal, developmental and cell proliferative  
 disorders -  
 XX  
 PS Claim 9: Page 191: 229pp: English.  
 CC AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human  
 CC transmembrane proteins designated HTPN-1 to HTPN-79, respectively.  
 CC The transmembrane protein have immunospecific, antiproliferative and  
 CC neuroprotective activities. The human transmembrane proteins,  
 CC polynucleotides encoding them and other compositions and methods from  
 CC the present invention, can be used for the diagnosis, treatment or  
 CC prevention of immune, reproductive, smooth muscle, neurological,  
 CC gastrointestinal, developmental and cell proliferative disorders. The  
 CC HTPN's can be used to treat or prevent disorders associated with a  
 CC decreased expression or activity of HTPN.  
 CC  
 XX  
 SQ Sequence 2329 BP: 674 A; 691 C; 569 G; 382 T; 13 other:

Query Match 99.4%; Score 2316; DB 21; Length 2329;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGATCTAGCATGTGTGGGAGAGGCGCCCTGGGAGCGAGTTGAGAAATCTGCCA 60  
 DB 1 GCAGGATCTAGCATGTGTGGGAGAGGCGCCCTGGGAGCGAGTTGAGAAATCTGCCA 60  
 QY 61 GTAATGGGGATGAGTACGAACTTCCCAATGTCTACACACAGGCGCCGCAACCC 120  
 DB 61 GTAATGGGGATGAGTACGAACTTCCCAATGTCTACACACAGGCGCCGCAACCC 120  
 QY 121 CAGCCAGCAACGCGACGACCAACCCCGCCCGCCAGAGACCTTCAACCTTAACAG 180  
 DB 121 CAGCCAGCAACGCGACGACCAACCCCGCCCGCCAGAGACCTTCAACCTTAACAG 180  
 QY 181 CCCAAGGAGAGCAACCAACTGCAATACCTCTCAGATGGTGTCAAGACATATGG 240  
 DB 181 CCCAAGGAGAGCAACCAACTGCAATACCTCTCAGATGGTGTCAAGACATATGG 240  
 QY 241 AAACACAGTTGGCATGCGCTTTCACAGACCTGTGGATGCCCTCAAGCTGAACCTCC 300  
 DB 241 AAACACAGTTGGCATGCGCTTTCACAGACCTGTGGATGCCCTCAAGCTGAACCTCC 300  
 QY 301 GATTACTATAGATCTATTAACGCGCTTATGATATGGAAACAATAAAGCGCTTGGAA 360  
 DB 301 GATTACTATAGATCTATTAACGCGCTTATGATATGGAAACAATAAAGCGCTTGGAA 360  
 QY 361 AACACTATTAAGTGAATGCTCAGAAATGATCAAGACTTCAACTATGTTTAAACAT 420  
 DB 361 AACACTATTAAGTGAATGCTCAGAAATGATCAAGACTTCAACTATGTTTAAACAT 420  
 QY 421 TGTTCATCTACAAAGCGCTGAGATGACATAGTCTTAATGAGCAAGCTGTGAAAG 480  
 DB 421 TGTTCATCTACAAAGCGCTGAGATGACATAGTCTTAATGAGCAAGCTGTGAAAG 480  
 QY 481 CTCCTCTGCAAAAAAATAATAGAGTACCCAGAGAAACCGAGATCATGATGTCAG 540  
 DB 481 CTCCTCTGCAAAAAAATAATAGAGTACCCAGAGAAACCGAGATCATGATGTCAG 540  
 QY 541 GCAAAAGGAGAGAGCTGGGAGAAAGAAACAGGAGCAAAACCTGCGCTTTCACAG 600  
 DB 541 GCAAAAGGAGAGAGCTGGGAGAAAGAAACAGGAGCAAAACCTGCGCTTTCACAG 600  
 QY 601 GTACCAAAACAGACTCAGATGCTCTGCGAGAGCCAGCCCTGAGCGGAATCC 660  
 DB 601 GTACCAAAACAGACTCAGATGCTCTGCGAGAGCCAGCCCTGAGCGGAATCC 660

QY 661 CTTCTGTGAGGCAAGGCTCAACCCCTTCTCTGCGTCAACCCGAGCTCATGCTCCAG 720  
 DB 661 CTTCTGTGAGGCAAGGCTCAACCCCTTCTCTGCGTCAACCCGAGCTCATGCTCCAG 720  
 QY 721 ACCCTGTATGACAGTGTGCTCCCAAGCACTGACAGCGCCCGCCAGTGCCCC 780  
 DB 721 ACCCTGTATGACAGTGTGCTCCCAAGCACTGACAGCGCCCGCCAGTGCCCC 780  
 QY 781 CAGCACAACCCCAACCGCTCCAGTCCCAAGCTCCCAAGCTCCCAAGCTCCCAAG 840  
 DB 781 CAGCACAACCCCAACCGCTCCAGTCCCAAGCTCCCAAGCTCCCAAGCTCCCAAG 840  
 QY 841 GCGGCAACCCCAACCGCTCCAGTCCCAAGCTCCCAAGCTCCCAAGCTCCCAAG 900  
 DB 841 GCGGCAACCCCAACCGCTCCAGTCCCAAGCTCCCAAGCTCCCAAGCTCCCAAG 900  
 QY 901 ACCCCCAACCACTTGAACCCCAATTCACAGACCTCTGCTGCGCCCGGAGCCAAAG 960  
 DB 901 ACCCCCAACCACTTGAACCCCAATTCACAGACCTCTGCTGCGCCCGGAGCCAAAG 960  
 QY 961 ACCAGCTGGGCGAGGCGGAGAGAGAGCGGCTGTGAACCTCCAAAGAGAGAG 1020  
 DB 961 ACCAGCTGGGCGAGGCGGAGAGAGAGCGGCTGTGAACCTCCAAAGAGAGAG 1020  
 QY 1021 CCGAGCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
 DB 1021 CCGAGCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
 QY 1081 TGTGTGAGGAGATCTCTCAAGAGATGTTTCCCAAGAGAGAGAGAGAGAGAG 1140  
 DB 1081 TGTGTGAGGAGATCTCTCAAGAGATGTTTCCCAAGAGAGAGAGAGAGAGAG 1140  
 QY 1141 TTCTAAGAGCTGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 DB 1141 TTCTAAGAGCTGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 QY 1201 CACCCATGAGATGAGCAACCAATCAAGTAACTGAGAGCGCGTGAATACGTTGCT 1260  
 DB 1201 CACCCATGAGATGAGCAACCAATCAAGTAACTGAGAGCGCGTGAATACGTTGCT 1260  
 QY 1261 CAGGATTTGTGCTGACAGCTCCGATGATGTTTCCCAAGTGAATGAGAGAGAG 1320  
 DB 1261 CAGGATTTGTGCTGACAGCTCCGATGATGTTTCCCAAGTGAATGAGAGAGAG 1320  
 QY 1321 GACCATGAGTGTGCTGACAGCTCCGATGATGTTTCCCAAGTGAATGAGAGAGAG 1380  
 DB 1321 GACCATGAGTGTGCTGACAGCTCCGATGATGTTTCCCAAGTGAATGAGAGAGAG 1380  
 QY 1381 AAGATCCGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1381 AAGATCCGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 CCCACAAAGTGTGAG 1500  
 DB 1441 CCCACAAAGTGTGAG 1500  
 QY 1501 AGTGAAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
 DB 1501 AGTGAAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
 QY 1561 GAGCAGCTCAAAAGCGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 DB 1561 GAGCAGCTCAAAAGCGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 QY 1621 CCAAAAGGAG 1680  
 DB 1621 CCAAAAGGAG 1680  
 QY 1681 GAAAGTGAAG 1740  
 DB 1681 GAAAGTGAAG 1740

QY 1741 AATTAATGACCAACGATATGTAGCAGAGAGCCAGCCCGCCATGAGAGCAAGCC 1800  
DB 1741 AATTAATGACCAACGATATGTAGCAGAGAGCCAGCCCGCCATGAGAGCAAGCC 1800  
QY 1801 CCTCCACGATGTAGTGGAGAGAGGACAAAGTCAAGCTATGTCTATGAGAGAGAG 1860  
DB 1801 CCTCCACGATGTAGTGGAGAGAGGACAAAGTCAAGCTATGTCTATGAGAGAGAG 1860  
QY 1861 CGGACGCTCAGCTTGACATCAACAGCTCCCGGCGAGAGCTGGCGCGGTGGTGCAC 1920  
DB 1861 CGGACGCTCAGCTTGACATCAACAGCTCCCGGCGAGAGCTGGCGCGGTGGTGCAC 1920  
QY 1921 ATCATTCAGTCAGCGGAGCCCTCCCTGAGAAATTCACACCCGACGAGATTGAATCGAC 1980  
DB 1921 ATCATTCAGTCAGCGGAGCCCTCCCTGAGAAATTCACACCCGACGAGATTGAATCGAC 1980  
QY 1981 TTTGAGACCCCTGAGACCCGCTCCACACTGCTGATGCTGGAGCGCTATGCTCTCTTT 2040  
DB 1981 TTTGAGACCCCTGAGACCCGCTCCACACTGCTGATGCTGGAGCGCTATGCTCTCTTT 2040  
QY 2041 GCGGAG 2100  
DB 2041 GCGGAG 2100  
QY 2101 NATGAG 2160  
DB 2101 NATGAG 2160  
QY 2161 TTTTACAG 2220  
DB 2161 TTTTACAG 2220  
QY 2221 AAACCGAG 2280  
DB 2221 AAACCGAG 2280  
QY 2281 CCTTTTCCCCAG 2329  
DB 2281 CCTTTTCCCCAG 2329

RESULT 2  
AAS03029 standard; cDNA; 667 BP.  
XX  
AC AAS03029;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Human diagnostic and therapeutic (dthp) cDNA sequence #18.  
XX  
KW Human diagnostic and therapeutic molecule; dthp; gene therapy;  
KW thalassemia; cardiovascular disorder; cell proliferative disorder;  
KW cancer; neurodegenerative disorder; autoimmune disorder;  
KW infectious disorder; inflammatory disorder; developmental disorder;  
KW Incyte ID number 2432679dec; intracellular signalling molecule; ss.  
XX  
OS Homo sapiens.  
XX  
PN M0200121836-A2.  
XX  
PD 29-MAR-2001.  
XX  
PF 19-SEP-2000; 2000MO-US25643.  
XX  
PR 23-SEP-1999; 99US-0155760.  
PR 24-SEP-1999; 99US-0155939.  
PR 24-SEP-1999; 99US-0156294.  
PR 28-SEP-1999; 99US-0156565.  
PR 28-SEP-1999; 99US-0156624.  
PR 28-SEP-1999; 99US-0156625.  
PR 24-NOV-1999; 99US-0167410.  
PR 24-NOV-1999; 99US-0167453.

PR 24-NOV-1999; 99US-0167517.  
PR 24-NOV-1999; 99US-0167520.  
PR 24-NOV-1999; 99US-0167542.  
PR 29-NOV-1999; 99US-0167943.  
PR 29-NOV-1999; 99US-0167945.  
PR 30-NOV-1999; 99US-0168197.  
PR 30-NOV-1999; 99US-0168265.  
PR 30-NOV-1999; 99US-0168429.  
PR 30-NOV-1999; 99US-0168432.  
PR 01-DEC-1999; 99US-0168468.  
PR 01-DEC-1999; 99US-0168599.  
XX  
XX (INCYTE GENOMICS INC.  
PA Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banyille SC;  
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;  
PI Hillman JL, Jones AL, Yu YJ, Greenwalt LB, Panzer SR;  
PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK;  
PI Amshay S, Fong WT;  
XX  
DR WPI; 2001-281607/29.  
XX  
XX Novel diagnostic and therapeutic polynucleotides, used in disease  
PT diagnosis and for gene therapy of conditions such as cancer and  
PT thalassemia  
PT  
PS Claim 1; Page 263; 299pp; English.  
XX  
XX The present sequence for human diagnostic and therapeutic (dthp) cDNA  
CC sequence #18 is 1 of 71 (AAS03012-AAS03082) novel sequences described  
CC in the invention. The present sequence (Incyte ID No: 2432679dec)  
CC encodes an intracellular signalling molecule. The dthp polynucleotides  
CC may be used to diagnose a condition disease or disorder associated with  
CC human molecules. They can be used to identify the presence of similar  
CC nucleic acids. Dthp polynucleotides may be used to generate hybridisation  
CC probes for use in chromosomal mapping. Polypeptides (Dthp) encoded by  
CC dthp are used to screen for molecules which bind to them and modulate  
CC their activity. Dthp polynucleotides can be used for gene therapy of  
CC disorders such as severe combined immunodeficiency syndrome (SCID),  
CC cystic fibrosis, thalassemia, haemophilia resulting from Factor VIII  
CC or IX deficiencies, cardiovascular disorders e.g familial  
CC hypercholesterolaemia (FH), cell proliferative disorders e.g. cancers,  
CC neurodegenerative disorders, autoimmune/inflammatory disorders,  
CC infectious disorders and developmental disorders. The antibodies can be  
CC used to analyse protein expression levels.  
XX  
SQ Sequence 667 BP; 229 A; 157 C; 149 G; 132 T; 0 other;  
Query Match 24.0%; Score 559.4; DB 22; Length 667;  
Best Local Similarity 98.8%; Pred. No. 1.8e-115;  
Matches 574; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 4 GGGATCACTACATGCTGCGGAGAGGCGCTT-GGGACGAGATTGAGAAATTCGCCAGT 62  
DB 10 GGGATCACTACATGCTGCGGAGAGGCGCTTGGGCGGAGATTGAGAAATTCGCCAGT 69  
QY 63 AATGGGGATGAGCTAGAGAACTTCCCAATGTCTACAAACAGAGGCCGAGCCCAACCCCA 122  
DB 70 AATGGGGATGAGCTAGAGAACTTCCCAATGTCTACAAACAGAGGCCGAGCCCAACCCCA 129  
QY 123 GCCAGCCAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 182  
DB 130 GCCAGCCAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 189  
QY 183 CAAGAGCAGACCAACCACTGCAATCTCTAGAGTGTGCTCAAGACACTATGGA 242  
DB 190 CAAGAGCAGACCAACCACTGCAATCTCTAGAGTGTGCTCAAGACACTATGGA 249  
QY 243 ACACAGCTTTCATGCGCTTTCAGAGAGCCTGTGATGCGCTCAAGCTGAACTCCCTGA 302  
DB 250 ACACAGCTTTCATGCGCTTTCAGAGAGCCTGTGATGCGCTCAAGCTGAACTCCCTGA 309  
QY 303 TTAATAAGATCAATTAAGAGCGCTATGATGAGAGAAATTAAGAGCGCTTGAGAAA 362

|||||  
Db 310 TTACATATAGATCATTTAAACCCCTATGATATGGAGAAATAAAGAACCGCTTGGAAAA 369  
Oy 363 CAACATATCTGGAATGCTCAGGAATGTATCCAGACTTCAACACATGTTTCAAAATG 422  
Db 370 CAACATATCTGGAATGCTCAGGAATGTATCCAGACTTCAACACATGTTTCAAAATG 429  
Oy 423 TTACATATCTGGAATGCTCAGGAATGTATCCAGACTTCAACACATGTTTCAAAATG 482  
Db 430 TTACATATCTGGAATGCTCAGGAATGTATCCAGACTTCAACACATGTTTCAAAATG 489  
Oy 483 CTCTTGGCAAAAAATAAATGAGCTACCCACAGAAAGAACCGAGATCATGATCCAGGC 542  
Db 490 CTCTTGGCAAAAAATAAATGAGCTACCCACAGAAAGAACCGAGATCATGATCCAGGC 549  
Oy 543 AAAAGGAAAGAGACGTGGAGGAAAGAAACAGGACAGCA 583  
Db 550 AAAAGGAAAGAGACGTGGAGGAAAGAAAGAGGTTATCA 590

RESULT 3  
ABK83660  
ID ABK83660 standard; cDNA; 4664 BP.  
XX  
AC ABK83660;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #231.  
XX  
XX Human, ss; granulocytic cell; DNA chip; bacterial infection;  
KM viral infection; parasitic infection; protozoal infection;  
KM fungal infection; sterile inflammatory disease; psoriasis;  
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KM cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KM adult respiratory distress syndrome; inflammatory bowel disease;  
KM Crohn's disease; ulcerative colitis; periodontal disease;  
KM granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
XX  
PR 03-OCT-2000; 2000US-237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
DR WPI; 2002-435328/46.  
XX  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity  
XX  
PS Claim 1; SEQ ID No 231; 114pp; English.

The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by CC DNA chip analysis as given in the specification, and comparing CC the expression level to an expression level in an unactivated CC GC, where differential expression of Gs is indicative of GCA. CC Also included are modulating (M2) GA by contacting GC with an agent CC that alters the expression of at least one gene in Gs; (2) screening (M3) CC for an agent capable of modulating GCA or an inflammation (especially CC chronic) in a tissue, an allergic response in a subject, exposure of a CC subject to a pathogen or sterile inflammatory disease using the CC gene expression profile; (3) detecting (M4) an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a CC subject to a pathogen or sterile inflammatory disease, by detecting the CC level of expression in a sample of the tissue of gene(s) from Gs, where CC the level of expression of the gene is indicative of inflammation; CC (4) treating (M5) an inflammation (especially chronic) or in a tissue, CC an allergic response in a subject, exposure of a subject to a pathogen CC or sterile inflammatory disease, by contacting a tissue having CC inflammation with an agent that modulates the expression of gene(s) CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for CC modulating GA; M3 is useful for screening an agent capable of modulating CC GCA preferably in an inflammation in a tissue; M4 is useful for CC detecting an inflammation (especially chronic) in a tissue, an allergic CC response in a subject, exposure of a subject to a pathogen or sterile CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal CC reperfusion injury, ARDS, adult respiratory distress syndrome, CC inflammatory bowel disease, Crohn's disease, ulcerative colitis, CC periodontal disease, also bacterial infection, viral infection, CC parasitic infection, protozoal infection, fungal infection and M5 is CC useful for treating one of the above conditions. The present CC sequence represents a gene differentially expressed in granulocytes. CC Note: The sequence data for this patent did not form part CC of the printed specification, but was obtained in electronic CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 4664 BP; 1084 A; 1336 C; 1293 G; 951 T; 0 other;

Query Match 19.8%; Score 460.2; DB 24; Length 4664;  
Best Local Similarity 57.0%; Pred. No. 5.9e-93;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;

Oy 143 CCNACCCCCGCCGCCGAGACCTCCACCCCTAACACGCCCAAGAGGACCAACCAAC 202  
Db 1877 CCNACCCACACCCGCCGCCGAGGAGTGCTCAATCCCAAAAGCAGAGGAGTACCAACACAC 1936  
Oy 203 TGCATATCTGCTCAGAGTGTGCTCAAGACACTATGAAACCAAGTTCATGCGCTT 262  
Db 1937 TGCATATCTGCTCAGAGTGTGCTCAAGACACTATGAAACCAAGTTCATGCGCTT 1996  
Oy 263 TTCAGGACGCTGTGATGCTGCGTCAACGCTGACCTCCCTATGATCATGATTAATAA 322  
Db 1997 TTCGCGACGCTGTGATGCTGCTCAACGCTGCTTACCGGATTTATCAAAATTAATAAAC 2056  
Oy 323 CGCCTATGATATGGAACAAATTAAGAACGCTTGGAAACCAATTAATCTGGAATGCTC 382  
Db 2057 AGCCTATGACATGGGTACTATTAAGAGAGACTTGAATAACAATTTATTTGGCGCTT 2116  
Oy 383 AGGATGTATCCAGAGCTTCAACACTATGTTTCAAAATGTTTACATCTCAACACAGCTG 442  
Db 2117 CAGAGTGTATGCAAGATTTTAATACCATGTTCACCAAGTTCATTAACACCAACCCCA 2176  
Oy 443 GAGATGACATATGCTTAATGGCGAAGCTGTGAAAGAGCTCTCTCTCAAAATAATTAAG 502  
Db 2177 CTGATATATTTGCTTAATGGCGAAGCTGTGAAAGAGCTCTCTCTCAAAATAATTAAG 2236  
Oy 503 AGCTACCAAGAAAGAAACGAGATCATGATGCTCAGGCAAAAGAGAGAGACGTTGGA 562  
Db 2237 CAATGCCACAGAAAGAAAGAGAGCTGTAGTGACATCCCTTAAGACAGCACAAGAG 2296  
Oy 563 GGAAGAAACAGGAGACAGCAAAACCTGGCGTTTCCAGGTTACCAACACACTCAAGCAT 622  
Db 2297 GGGCCCAAGTTGGAGAGCTTCAGGGGAGGTTTACAGTCCCATCAGTGGCTCGCTCT 2356  
Oy 623 CGAGCTCTCCGAGAGCCGAGACCCCTCAGCAAGATCCCTCTCTGAGAGGAGCCCTC 682  
Db 2357 -----CTTCTGTGTACACACAGCCCTGTATATCTCTCACCCTGAGATACCTACACG 2410  
Oy 683 ACCCTTTCCTGCGGACACCCGAGACCTCATGCTCAGACAGCCCTGATGACAGTGTGTC 742  
Db 2411 TCTGCAACATTCGCCACCATCATGATATTC-----T 2443  
Oy 743 CTCCCGAGCACTGACAGAGCCGCCGAGTGCCTCCGACGACCAACCCGACCGCTC 802





CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 4664 BP; 1084 A; 1336 C; 1293 G; 951 T; 0 other;

Query Match 19.8%; Score 460.2; DB 24; Length 4664;  
Best Local Similarity 57.0%; Pred. No. 5.9e-93;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;

```
OY 143 CCAACCCCGCCCGCCGAGACCTCCACCCCTAACAAGCCCAAGAGAGACCAACCAAC 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1877 CCAACCCCGCCCGCCGAGAGTGTCCATCCCAAAAAGAGAGAGATTACCAACAGC 1936
OY 203 TCGAATACCTGCTCAGAGTGTGCTCAAGACACTATGGAACCCAGTTTGCGTTC 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1937 TCGAATACCTACCAAGAGTGTGATGATGAGGCTCTGTGGAACATCATGTCGATG 1996
OY 263 TCCAGAGCCTGTGATGTGCGTCAAGCTGACCTCCCTGTTACTATTAAGATATTA 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1997 TCCGAGAGCCTGTGATGTGCTCAAACTGGGTCTACCGGATTAATCAAAAATTATA 2056
OY 323 CGCCTATGATATGGAACATTAAGAGCGCTTGGAAAAACAATACTATGATGATGCT 382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2057 AGCTTATGACATGGTACTATTAAGAGAGACTTGAACAACTTATATGCGGCTCT 2116
OY 383 AGAATGTATCCAGACTTCAACATATGTTTACAATTTGTTACATCTACACAGCCTG 442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2117 CAGAGGTATGCAAGATTTTAAATACCATGTTCACCAACTGTACATTTTACAAC 2176
OY 443 GAGATGATGATCTTAATGAGGAGAGCTGTGAAAAAGCTCTCTGCAAAAATAAT 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2177 CTGATATATTTGTCTTAATGTGCAAAACGCTGAAAAAGATTTCTTACAGAAAGT 2236
OY 503 AGCTACCAACAGAAAGAACCGAGATCATGATATGTCAGGCAAAAAGAGAGAGAG 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2237 CATATGCCAAGAAAGAAAGAGCTGTAGTACCATCCCTTAAGAAAGCAACAAG 2296
OY 563 GGAAGAAACAGGAGACAGCAAAACCTGGGTTTCCAGGTTACCAACACATCTCA 622
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2297 GGGCCAAAGTTGGCAGGCTCCAGGAGAGTGTACAGTGCCTCATCGTCCGCTC 2356
OY 623 CCACTCTCCGAGACCCCTCAGCCGAACTCTCTCTGTGAGGAGCCAGCCTC 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2357 CTTCTGTGTACACACAGCCCTGTATCTCTCTCAGCTGAGATTAACCTAC 2410
OY 683 ACCCTTCCCTGCGTACCCCGGACCTCATGCTCCAGACCCCTGTATGACAGTGTG 742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2411 TCTTCAACATTCCTCCACCATCATGCTATTTCC-----T 2443
OY 743 CTCCCAAGCACTGACAGAGCCCGCCGAGTGCCTCCCAAGCCCAACCCCAACCCGCT 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2444 CTCCCACTTCTCAAGCTCTTCCAGCTGTGAGACCCCGCTCTCTGCTGTTAAGCA 2503
OY 803 CAGCTCTCCAGCCGCTGACAGAGCCACCACTCATGCGGAGCCACCCACAGCTGTG 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2504 CTCCAGCCCAAGCCCTTGGCAAGAAAAAGGCT----- 2537
OY 863 AACAAGAGAGAGAGTGAAGAGAAAGACACCAACCAACCCCAACCCCAACCTTGA 922
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2538 ---AAAGGAGAAAGAGATACCAACCCCTACCTAAGCACTTGTGGCTCTG 2593
OY 923 TTTACAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2594 CTCCCACTGAGCCCTCTCTGAGAGCTTGAAGGACACAGGCTTCCCTCTATCTG 2653
OY 980 GGGAGAGAGAGAGAGCTGTAAGAACTCTCAAGAAAGAGAGAGAGAGAGAGAGAG 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2654 GAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2713
OY 1040 CAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2714 ACCAGAGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2773
```

```
OY 1100 AGAGATGTTTGGCAAGAAAGACCCGCTACGCTGCGCTTCTACAAAGCTGTGAGC 1159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2774 AGAGATGTTTGGCAAGAAAGACCCGCTACGCTGCGCTTCTACAAAGCTGTGAGC 2833
OY 1160 TGGAGGACCTGAGGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2834 CTCTGACCTGAGGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2893
OY 1220 CAATCAAGTCTTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2894 CTCTGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2953
OY 1280 TCCGATGATGTTTCTCAACTGCTATTAAGTACCAACCTCTGACCTGAGAGTGTG 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2954 TACGCTTATGTTCTCAACTGCTATTAAGTACCAACCTCTGAGAGAGAGAGAG 3013
OY 1340 TGGCCGCAAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3014 TGGCAGAGAAAGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3073
OY 1400 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3074 TGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3133
OY 1457 CCGCCGCTTCACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3134 CAGAGTCTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3193
OY 1497 -----GAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3194 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3253
OY 1541 AGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3254 ATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3313
OY 1601 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3314 CCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3373
OY 1660 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3374 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3433
OY 1706 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3434 CAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3493
OY 1750 -----AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3494 CTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3553
OY 1772 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3554 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3613
OY 1832 AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3614 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3673
OY 1892 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3674 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3733
OY 1952 ATTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3734 ATTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3793
OY 2012 AGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3794 AGCTT--GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3831
```

XX	AAH18591	standard; cDNA; 1884 BP.
ID	AAH18591	
XX	AAH18591;	
XX	26-JUN-2001	(first entry)
XX	Human CDNA sequence SEQ ID NO:18782.	
DE	Human CDNA sequence SEQ ID NO:18782.	
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
KW	Homo sapiens.	
OS	Homo sapiens.	
XX	EPI074617-A2.	
PN	07-FEB-2001.	
PD	28-JUL-2000; 2000EP-0116126.	
PF	29-JUL-1999; 99JP-0248036.	
XX	27-AUG-1999; 99JP-0300253.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	02-MAY-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX	(HELI-) HELIX RES INST.	
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX	WPI, 2001-318749/34.	
DR	Primer sets for synthesizing polynucleotides, particularly the 5602	
XX	full-length cDNAs defined in the specification, and for the detection	
PT	and/or diagnosis of the abnormality of the proteins encoded by the	
PT	full-length cDNAs -	
XX	Claim 8; SEQ ID 18782; 2537bp + CD ROM; English.	
PS	The present invention describes primer sets for synthesising 5602	
XX	full-length cDNAs defined in the specification. Where a primer set	
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
CC	to the complementary strand of a polynucleotide which comprises one of	
CC	the 5602 nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in	
CC	the specification. The primer sets can be used in antisense therapy and	
CC	in gene therapy. The primers are useful for synthesising polynucleotides,	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH016 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
CC	AAB93883 represent human amino acid sequences; and AAH13629 to AAH13632	
CC	represent oligonucleotides, all of which are used in the exemplification	
CC	of the present invention.	
XX	Sequence 1884 BP; 492 A; 536 C; 451 G; 405 T; 0 other:	
SQ		
	Query Match	15.9%; Score 371; DB 22; Length 1884;
	Best Local Similarity	58.3%; Pred. No. 3.9e-73;
	Matches 820; Conservative	0; Mismatches 495; Indels 92; Gaps 6
0Y	156 CCCAAGACCTCCACCCTCAACAAGGCCGAGCGACCAACCAACTGCATTACTGCT	215
DB	422 CCGGAGGTGTCATCCCAAAAAGCAGACGAGTTACCAACCACTTCCAATTACTTAC	481

OY	216	CAGATGGTGGTCAAGACACTATGGAAGCAACCGAGTTGATGGCCCTTTCACAGACCTCT	275
Db	482	CAGGTAATGATGAAGGCTCTGTGGAAACATCGATTGCGATGGCCATTCGCGACACCTGT	541
OY	276	GGATGGCCCTCAAGCTGGAACCTCCCTGATTTCTATTAAGATCATTTAAAGCCCTATGGATAT	335
Db	542	GGATGCTGTCAAACTGGGGTCTACCGGATTTATCCAAAATTTATTAACAGCCCTATGGACAT	601
OY	336	GGGAACATTAAGAAAGCGCTGGAAAAACAATACTATTACTGGAAATGCTCAGGAATGTATCCA	395
Db	602	GGGTACTATTAAAGAGGAGACTTGAAAAATTTATTTATGGGCGCTTCAGAGTATATGA	661
OY	386	GGACTTCAACACTATGTTTACAAATTTTTCATCTTCACACACAGCCGTGAGATGACATAGT	455
Db	662	AGATTTTAAATCCATGTTTCCAACTGTTTCATTTCATTAACAACGCCCACTGAATATATGT	721
OY	456	CTTATGTCAGACACTCTGTGAAAAAGCTCTTGTGCAAAAAATTAATGACTCTACCCACGA	515
Db	722	CTTATGTGCACAAACGCCCGAAAAAGATATTTCTTACAGAAAGTTTGATCATATGCCACAA	781
OY	516	AGAAACCGAGTTCATGATAGTTCAGAGCAAAAGGAAGAGACGTGGGAGAAAAAGAAACAG	575
Db	782	AGAACAAAGCTGTGATGACATCCTTAAGAACAGCACAAGAAAGGGGCCAAGTGTGGC	841
OY	576	GACAGCAAAACCTGGCTTTCCAGAGGTACCAACACAACTCAAGCATGCGACTCTTCGCCA	635
Db	842	AGCGCTCCAGGCGAGTGTTCACAGTGGCCATTCAGGTGCTGCGCTCTTCTTCTG-----T	895
OY	636	GACCCAGACCCCTTAGCCGGAATCCTCCTGTGACAGGCAAGCCACACCCCTTCCCTGCG	695
Db	886	GTCACACACAGCCCTGTATCTCTCCACTGTGAAATCTTACACACTGTCTTAACATATCC	955
OY	696	CGTCAACCCCGAGCTCATGTGTCCAGACCCCTGTATGACAGTGTGCTCCCAACCACT	755
Db	956	CCACCCATCAAGTATTTTC-----TCTCCACTTCTCA	988
OY	756	GCAGAGCGCCCCCGCCAGTGGCCCCCCCCACGCCACAACCCCAACCCGCTCAGCTCCCGACG	815
Db	989	GTCCTTGTACTCTGTGGACCCCCCGCTCTTGTGTACTGAGAGCTCTCCAGCCAGCC	1048
OY	816	CGTACAGAGCCACCCACCCATTCATCGGGGCCACCCCAACGCTGTGAATACAAAGAAAGG	875
Db	1049	CCT-----TCCAAAGAAAAAAGG	1067
OY	876	AGTAGAGNGAAGACAGACACACACACCCCCACACATTTGACCCCAT-----	923
Db	1068	CGTAAAGGGAAAGACATATCTACACACCCCTTAACCTACAGCAGATTTGGCTCGCTG	1127
OY	924	TCACGAGACCCCTCGCTGCCCCCGAGACCCCAAGACCAACAGCTGGGCC--AGCGGCG	980
Db	1128	TCCAGCTAGCCCTCTCTGGGAGCTTGTAGCCCTTAAGGACAGCAGGCTCTCCCTATCTAG	1187
OY	981	GGAGAGCAGCGCGGCTGTGAAACCTCCAAAGAAAGAGCGGCCGAGCTCCACAGACACC	1040
Db	1188	AGAAGATGTGTGCCCATTCAGAGCCCCCAGCCAAAGACTTGCCTGAGCTCCACCAACA	1247
OY	1041	AGCACAGAGAAAGCAGCAAGGTCTTCGAGACGCTCAAGTGTGACGAGCGATCACTCA	1100
Db	1248	CCAGAGCTTAAGAAAGAAAGAGTTTCAGAACAGTTAAACAATGCAATGGCATTTTGA	1307
OY	1101	GGAATGTTTGGCCAAAGACAGCGCGGCTTACGCTGCGCCCTTCTACAGCCCTGTGGAGCT	1160
Db	1308	GGAATGTTTCTCTTAAGAAAGATGCTGCTGTGTTGGCTTCTTAACCAAGTGGATGCG	1367
OY	1161	GGAGGCACTGGGCTTACAGCATACGTGACATCAATCAAGCAACCCCATGGACCTTAGAGC	1220
Db	1368	TTTCTGCACTTGGCTTGCATACATCAATGACATCAATTAAGCAACCCCATGGACCTTAGAGC	1427
OY	1221	AATCAAGCTTAAACTGGAGGCGCGTGAATACCGTGAATGCTCAGAGATTGTGGTCTGACGT	1280
Db	1428	TGTCAAGGGAAGATGGAAGAACGTAATTAACGGGATGACAGAGAGTTTGGCTGCTGAATG	1487
OY	1281	CCGATGATGTTTCCAACTGCTATAGTACAACCTTCCTGACATGAGGTGTGGCCAT	1340



Db 1012 GCTTTGGAGCTCCATTAATGATGAGCTTGTCAAAAATCCGAGGATCTGGAACTATT 1071  
 QY 1225 AAGTCTAACTGAGGCCCTGAGTACCGTATGCTAGAGTTGGTGTGCTGACCTCGCA 1284  
 Db 1072 AAGGAGAAAATGAGTAAACCAAGATATAGATGATCTATTTGGCGAGATGTAGA 1131  
 QY 1285 TTGATGTTCTCAACCTGCTATTAAGTACAACTCTGACATGAGGTGGTGGCATGGCC 1344  
 Db 1132 TTATGTTTCATGAATGCTGACAACTACAACTCTGACATGAGGATTTGTGACATGGCA 1191  
 QY 1345 CGCAAGCTCCGAGATGCTTGAATGCGCTTCCCAAGATCCGCGAGACGCTGAGAG 1404  
 Db 1192 AGAATGCTCAGATGTTTGTGAAACGATTTTCAAAAGATCCGATTTGAACCTGTTAG 1251  
 QY 1405 CCAGTGTGCGCTGCTGCTCCCGGAGTGGCCCTCCCAAGAGTTGTGGCCCGCC 1464  
 Db 1252 AGTATGCTTTATTTATACATAAAACAGATATCACAGAAACCACTGCT----- 1299  
 QY 1465 TCATCCAGCAGCAGCAGCAGCAGTACCTCTGAGCAGTACAGTTCGATGATGATCT 1524  
 Db 1300 -----AGAGAGAACACTTAATGAAGCCCTCTGTGAAGGGAAC---TCTTGTGATGATCT 1350  
 QY 1525 GAGGAGAGCAGCAGCAGCAGCAGTACCTCTGAGCAGTACAGGAGCAGTCAAAAGCCTGACAG 1584  
 Db 1351 GAAATGATGAGCAGTGAAGCTGCTGCAAAAGCTTCAAGGAGCTTAAAGCTGATACATCA 1410  
 QY 1585 CAGCTTCAGCCCTCTCTCAGCCAGCAGCAAAACCAAGCAAAAGAGAGAAAGACAAG 1644  
 Db 1411 CAGCTCCAGGTTTGTGCTCCAGTACCTTTCCTGAAGCTAAATTAATAAG----- 1458  
 QY 1645 AAGGAAAAAGAAAAAGCAACAAAAAGAAAGAGAGAGTGAAGAGATTAATAAGCA 1704  
 Db 1459 -----AAAGAGAACTTAAAGAAAAAGAAAAAGAAAAAGAGTTAATAAGAGC 1506  
 QY 1705 AAACCAAGAGAACTCTCTCTTAAAGAGCAAGAAAAATATAGCAGACAGCAATGTG 1764  
 Db 1507 AATATAAA-----TCCAAAGAAAAATGTGTAGCAAAATGAGGCTTAAAGAAAAAGTCC 1557  
 QY 1765 AGCAAGAAAGAGCAGCAGCAGCAGCAAGAGCAAGCCCTCCAGCTATGATGAGAGAGAA 1824  
 Db 1558 AAGGAAATATCAGCAAAAGAAAAAGAAACAAACAGTTCATTTGCTTAAAA---TCTGAAGAT 1614  
 QY 1825 GAGGACAAAGTGCACCACTATGCTCTATGAGAGAGAGAGGAGCTGAGCTTGACATCAAC 1884  
 Db 1615 GAAGATATATGCTTAACCTATGATGATGAGAAAAAGGAGTTAAGCTGAATATAAAC 1674  
 QY 1885 AAGTCCCCCGGAGAGAGCTGGCGGTGTGACATCATCCAGTCAACGGAGAGCCCTCC 1944  
 Db 1675 AACTCTCCCTGAGATTAACCTTGGGCGAGTATTCACATTAATVACAATCAAGAGAGCTTCT 1734  
 QY 1945 CTGAAGAAATTCACACCCGAGAGATGAAATGCACTTGAAGACCCGAGCCGCA 2004  
 Db 1735 CTGAGCAATTCATCTCTGATGAGATGAGATAGATACCTTTGAACACCTGAAAGCATCAACA 1794  
 QY 2005 CTGCGTAGCTTGA 2019  
 Db 1795 CTAGAGAAATTAGAA 1809  
 RESULT 7  
 AAX40172  
 ID AAX40172 standard; DNA: 3106 BP.  
 AC AAX40172:  
 XX 02-JUL-1999 (first entry)  
 DT MO9904265 Seq ID No: 686.  
 DE  
 XX Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.

XX OS Homo sapiens.  
 XX PN MO9904265-A2.  
 XX PD 28-JAN-1999.  
 XX PE 15-JUL-1998; 98W0-US14679.  
 XX PR 22-JUN-1998; 98US-0102322.  
 XX PR 17-JUL-1997; 97US-0896164.  
 XX PR 10-OCT-1997; 97US-0061599.  
 XX PR 10-OCT-1997; 97US-0061765.  
 XX PR 11-OCT-1997; 97US-0948705.  
 XX PR 11-OCT-1997; 97GB-0021697.  
 XX PA (LUDW-) LUDWIG INST CANCER RES.  
 XX PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 XX PI Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;  
 XX PI Tureci O;  
 XX PR WPI: 1999-132448/11.  
 XX DR  
 XX PT New isolated cancer associated nucleic acids and polypeptides -  
 XX PT isolated using sera from cancer patients, used to develop products  
 XX PT for the diagnosis, monitoring or treatment of cancers  
 XX PS Claim 67; Page 730-731; 787pp; English.  
 XX CC The invention relates to a method for diagnosing a disorder characterised  
 XX CC by expression of a human cancer associated antigen precursor coded for by  
 XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 XX CC biological sample isolated from a subject with an agent that specifically  
 XX CC binds to the NAM, an expression product or a fragment of an expression  
 XX CC product complexed with an HLA molecule; and (b) determining the  
 XX CC interaction between the agent and the NAM or the expression product as a  
 XX CC determination of the disorder. The products and methods can be used in  
 XX CC the diagnosis, monitoring, research, or treatment of conditions  
 XX CC characterised by the expression of various cancer associated antigens.  
 XX CC The invention provides nucleic acid sequences and encoded polypeptides  
 XX CC which are cancer associated antigen precursors expressed in human breast  
 XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 XX CC lung cancer.  
 XX SO Sequence 3106 BP; 1179 A; 560 C; 594 G; 773 T; 0 other;  
 Query Match 12.6%; Score 294.2; DB 20; Length 3106;  
 Best Local Similarity 53.1%; Pred. No. 7.3e-56;  
 Matches 995; Conservative 0; Mismatches 673; Indels 207; Gaps 9;  
 QY 145 AACCCCGCGCCCGAGAGACCTCCAAACCTTAACAGCCCAAGAGCAGCAACCAACTG 204  
 Db 144 AACCTCTCCACCAAGATATATTAATTAAGAAAAATGGCGAGTTGACAAATCACTT 203  
 QY 205 CAATACCTGCTCAGAGTGTGCTCAAGACACTATGAGAAACACAGTTTGATGCTTTC 264  
 Db 204 CAGTATCTACAAAAAGTTGCTCTAAGAGATTATGAGAGCAATGTTTCAATGGCCCTTT 263  
 QY 265 CAGCAGCTGTGATGCCGCTCAAGCTGAACTCCCTGATTACTATAGACTATTAAGAG 324  
 Db 264 CAAGCTCCTGTGATGCTGTAAGAACTAAAGTGCCTGATTATTAACATTAATAAAAA 323  
 QY 325 CCTATGATATGGAACAATTAAGAGCGCTGGAAGAAACAACATATATAGGAATGCTAG 384  
 Db 324 CCAATGATTTTAAATACATTAATTAAGAGCGCTTGAGAAATTAATTAAGCGAGCTTCA 383  
 QY 385 GAATGTATCAGACTTCACACTATGTTTACAAATTTTACATCTACACAAAGCTTGA 444  
 Db 384 GAATGTATAGAAAGACTTCATACATATGTTTCAAAATTTTATTTATTAACAAAGCTTGA 443  
 QY 445 GATGACATAGTCTTATATGAGCAAGAGCTGTGAAAAAGCTCTTTCGCAAAAAATTAATAG 504

Db 444 GATGACATTGTTCTTATGGCACAAGCTCTAGAGACGTTTATGCGAAATTATCTCAG 503  
QY 505 CTACCCACAGAGAAACCGAGATCATGATAGTCCAGCGAAAGAGAGACGTTGGAG 564  
Db 504 ATGCCACAAGAG-----AGCAAGTTGGGTGTTAAGGAAGA 542  
QY 565 AAGAAACAGAGGACACCAAAACCTGGCGTTTCCACGGTACCAACACACTCAAGCTC 624  
Db 543 ATCAAAAGAGGACACTCAACGAAATATAGCTGTTCTTCTCTAAGAAATATCATCC 602  
QY 625 ACTCTCCGAGACCCAGACCCCTCAGCCGATCTCTCTCTGTCAGCGCCGCTCAG 684  
Db 603 AGCGCAACAGAAAAAGATTATTAAGCAGCAAGAAATCTCTCTGATTTCTTAAGACA--- 659  
QY 685 CCCCTCCCTGCGCTCAGCCCGGACCTCATGTCAGAACCCCTGTCATGACAGTGTGCT 744  
Db 660 -----TCTATT 659  
QY 745 CCCGACCCACTGCAGACGCCCGCCGAGTCCCCCGCCAGCACAAACCCCGCTCCA 804  
Db 660 -----TCTATT 665  
QY 805 GCTCCGACGCGGTACAGAGCCACCCATCATCGCGGCAACCCACAGCCTGTGAAG 864  
Db 666 TCTCCCTTGACCTGGTACAGGAGCTTCAGTCACTCCAGTTCAACAACTGGGCCCA 725  
QY 865 ACAAGAGAGGAGTGAAGAGAAAGACAGACACCCACCCACCATTTGACCCCTT 924  
Db 726 GTTACAAAGGTGTGAAGAGAAAGATCAACAACCTCTGCAAC--TTGACAGTTA 783  
QY 925 CAGAGACCCACCTCGCTGCGCCCGGAGCCCAACCAACCAAGCTGGGCCAGCGCGGAG 984  
Db 784 AAGCAAGTGTGATTTTCTCCAACTTCAACAAAAATCATGTCGACCTCATTTAA 843  
QY 985 AGCAGCGCGCTGTGAAACCTCCAAAGAGAGAGCTGCCGACTCTGACGACACCCAGCA 1044  
Db 844 AAGAA-----AATATGCCAAAGAAATGTTTGGCAATTTCTACGAAACAATTAAT 893  
QY 1045 CCAGAGAAAGAGCAGCAAGGCTCTGGAGCAGCTCAAGTGTGACGCGCATCTCAAGAG 1104  
Db 894 GTTGTGAGAGCTGTAAAGTACTGAACATTAAGGACGTGTAGTGTGATTTTAAAGAA 953  
QY 1105 AAGTTTCCCAAGAGACCGCGCTACGCTGCGCTCTACAAAGCTGTGGAGCTGTGAG 1164  
Db 954 AAGCTTGCAGAAAGAAATTTTATATGATGAGGCGCTTTATATCTCTGTGACGTTAAT 1013  
QY 1165 GCACCTGGCTTACAGCAGCTACTGTGACATCATCAGACCCCATGACATGACACATC 1224  
Db 1014 GCTTGGAGACTCCATACATCTATGACGTTGTCAAAAATCCGATGATCTTGGAACTATT 1073  
QY 1225 AAGCTTAACTGGAGCCCGGTAGTACCGTGTAGTCTCAGAGTTTGGTGTGACGTCGA 1284  
Db 1074 AAGGAGAAATATGATTAACGAAATATTAAGATGATTAATGATGATGCGCAGATGTAGA 1133  
QY 1285 TTGATGTTCCCAAGTGTATATGATCAACCCCTGACCATGAGTGTGGCGACCC 1344  
Db 1134 TTATATGATGATGATTGCTACAAAGTACAAATCTCCAGATACGAAAGTTGGAACAATGCA 1193  
QY 1345 CGCAAGCTCCAGGATGTGTTGCAAAATGCGCTTGGCAAGATGCGGACGAGCCTGAGAG 1404  
Db 1194 AGAATGCTTCAGAGATGTTTTCGAAAGCAGATTTTTCAAAGATGCCGATGAACTGTTGAG 1253  
QY 1405 CCAAGTGTGCGCTGTCTCCCGGAGTGCCTCCACCAAGGTTGTGGCCCGCC 1464  
Db 1254 AGTATGCTTTATGTATCATCAAAACAGATATCAGAAACCACTGCT----- 1301  
QY 1465 TCATCCAGGAGACGACGAGCGATAGCTCTCTGACAGAGTTCGATGATGACTCT 1524  
Db 1302 -----AGAGAGAACTATATGAGCCTCTCTGAAAGGGAAC---TCTTCTGATGATTC 1352  
QY 1525 GAGAGAGAGCAGGACCGAGGCTGTGAGCTCCAGAGAGAGCTCAAAAGCCGTCAGAG 1584  
Db 1353 GAAGATGACGAGCTTAAGCCTCTTTCGAAAGCTTCAGAGAGAGCTTAAAGCTGTACATCAA 1412

QY 1585 CAGCTTGCAGCCCTCTCTCAGCCCGCAGACAGAACAAACCAAGAAAAAGAGAACAAAG 1644  
Db 1413 CAGCTCCAGGTTTGTGCCAAGTACCTTCTCAGCTAATATATAAG----- 1460  
QY 1645 AAGGAAAAAGAAAAAGCACAAGAAAGAGAGAGTGCAGAGATTAATAAAAAAGC 1704  
Db 1461 -----AAGGAAAGTCTAAAAAGGAAAAAGAAAAAGGTTAATTAACAGC 1508  
QY 1705 AAGCCAAAGAACCTCTCTTAATAAGACGAGAAAAATATATGACGACACCAATGTG 1764  
Db 1509 AATGAAAA-----TCAAGAAAAATGTGTGAGCAATGAGCTTAAAGCAAAAGTCC 1559  
QY 1765 AGCAAGAGAGAGCCAGCGCCCATGAAGAGCAAGCCCTCCACGATGATGAGTGAGAA 1824  
Db 1560 AAGAGAAATACGCCAAAGAAAGAAACAAACATTCATTTGCTTAAAA---TTGAAGAT 1616  
QY 1825 GAGGACAAAGTCAAGCTTATGCTTATGAGAGAGAGCGAGCTGACGTTGACATCAAC 1884  
Db 1617 GAAGATTAATGCTTAACCTATGATGATGAGAGAAAGGACAGTTAAGTGAATATAAAC 1676  
QY 1885 AAGCTCCCGCGGAGAAAGCTGGCGCGGTGTGACATCATTCAGTCAAGGAGCGCTCC 1944  
Db 1677 AAGCTCCCTGAGATTAACCTTGGCGGAGTATGTTCACTAATTAACATCAAGAGACCTTCT 1736  
QY 1945 CTGAAGATTTCAACCCCGCAGAGATTGAATGACCTTTGAGACCTGAAAGCCGTCACA 2004  
Db 1737 CTGAGCAATTCCAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1796  
QY 2005 CTGCTGAGCTTGA 2019  
Db 1797 CTAAGAGAAATTAGAA 1811

RESULT 8  
ABV25740  
ID ABV25740 standard; cDNA; 1862 BP.  
XX  
AC ABV25740;  
XX  
PF 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 25731.  
XX  
XX Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-18319P.  
XX  
PR 16-MAR-2000; 2000US-189862P.  
XX  
PR 25-MAY-2000; 2000US-207454P.  
XX  
PR 09-JUN-2000; 2000US-211314P.  
XX  
PR 18-JUL-2000; 2000US-219007P.  
XX  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 5167-5168; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV2213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1862 BP; 502 A; 514 C; 483 G; 363 T; 0 other;

Query Match 9.8%; Score 227.8; DB 23; Length 1862;  
Best Local Similarity 58.6%; Pred. No. 4,4e-41;  
Matches 555; Conservative 0; Mismatches 312; Indels 80; Gaps 6;

QY 841 GCGGCCACCCACAGCCTGTGAAGACAAGAGGAGTGAAGAGAAAGCAGACACAC 900  
DB 389 GCTCTCCACCCACAGCCTGTGAAGACAAGAGGAGTGAAGAGAAAGCAGATCTAC 448  
QY 901 ACCCCACACACATGACCCCAT-----TCACGAGCACCCTGCGTCCGCCCG 948  
DB 449 ACCCTACACCTACAGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 508  
QY 949 GAGCCCAAGACAAGAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1005  
DB 509 GAGCCTAAGCAGACAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 568  
QY 1006 CCAAG 1065  
DB 569 CCAAG 628  
QY 1066 TCGAG 1125  
DB 629 TCAG 688  
QY 1126 GCTTACGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1185  
DB 689 GCTTACGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 748  
QY 1186 TGTGACATCATCAAG 1230  
DB 749 CATGACATCATCAAG 808  
QY 1231 -----AACTGAG 1280  
DB 809 TGTGAG 868  
QY 1281 CCGATTGATTTTCCCACTGCTAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340  
DB 869 ACCGCTATTTTCCCACTGCTAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 928  
QY 1341 GGGCCGGAAGCTCAGAGATGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1400  
DB 929 GGGACGGAAGCTCAGAGATGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 988  
QY 1401 GGAAGCAG 1457  
DB 989 AGAAGCAG 1048  
QY 1458 CCGGCTTATTCAG 1496  
DB 1049 AGAGTCTTCCAG 1108  
QY 1497 -----GGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1541

DB 1109 TGAG 1168  
QY 1542 GCGGCTGCTGAGTCCAG 1601  
DB 1169 TCGCTTACAG 1228  
QY 1602 TCAGCCCCAG 1660  
DB 1229 CCAGGCTTCATATTCAG 1288  
QY 1661 AAAAGCAG 1707  
DB 1289 GAAAGCAG 1335

RESULT 9  
AAS26606  
ID AAS26606 standard; cDNA; 424 BP.  
XX  
AC AAS26606;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, Seq ID 785.  
XX  
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;  
KW cytoskeletal; cardiac; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angioneurosis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO20015322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.

PR	18-AUG-2000	2000US -0256279
PR	22-AUG-2000	2000US -0256281
PR	22-AUG-2000	2000US -0256286
PR	23-AUG-2000	2000US -0227182
PR	23-AUG-2000	2000US -0227182
PR	30-SEP-2000	2000US -0228287
PR	01-SEP-2000	2000US -0228287
PR	01-SEP-2000	2000US -0229343
PR	01-SEP-2000	2000US -0229344
PR	01-SEP-2000	2000US -0229345
PR	05-SEP-2000	2000US -0228509
PR	05-SEP-2000	2000US -0228509
PR	06-SEP-2000	2000US -0230437
PR	06-SEP-2000	2000US -0230437
PR	08-SEP-2000	2000US -0230438
PR	08-SEP-2000	2000US -0231242
PR	08-SEP-2000	2000US -0231243
PR	08-SEP-2000	2000US -0231244
PR	08-SEP-2000	2000US -0231343
PR	08-SEP-2000	2000US -0231413
PR	08-SEP-2000	2000US -0233408
PR	08-SEP-2000	2000US -0233414
PR	08-SEP-2000	2000US -0233414
PR	08-SEP-2000	2000US -0233414
PR	12-SEP-2000	2000US -0233061
PR	12-SEP-2000	2000US -0233061
PR	14-SEP-2000	2000US -0233397
PR	14-SEP-2000	2000US -0233397
PR	14-SEP-2000	2000US -0233398
PR	14-SEP-2000	2000US -0233399
PR	14-SEP-2000	2000US -0233399
PR	14-SEP-2000	2000US -0233400
PR	14-SEP-2000	2000US -0233401
PR	14-SEP-2000	2000US -0233401
PR	14-SEP-2000	2000US -0233063
PR	14-SEP-2000	2000US -0233064
PR	14-SEP-2000	2000US -0233064
PR	14-SEP-2000	2000US -0233065
PR	21-SEP-2000	2000US -0233065
PR	21-SEP-2000	2000US -0233423
PR	21-SEP-2000	2000US -0234274
PR	25-SEP-2000	2000US -0234997
PR	25-SEP-2000	2000US -0234997
PR	25-SEP-2000	2000US -0234998
PR	25-SEP-2000	2000US -0235484
PR	27-SEP-2000	2000US -0235834
PR	27-SEP-2000	2000US -0235836
PR	29-SEP-2000	2000US -0236327
PR	29-SEP-2000	2000US -0236367
PR	29-SEP-2000	2000US -0236367
PR	29-SEP-2000	2000US -0236368
PR	29-SEP-2000	2000US -0236370
PR	02-OCT-2000	2000US -0236802
PR	02-OCT-2000	2000US -0237037
PR	02-OCT-2000	2000US -0237037
PR	02-OCT-2000	2000US -0237038
PR	02-OCT-2000	2000US -0237039
PR	02-OCT-2000	2000US -0237040
PR	13-OCT-2000	2000US -0239935
PR	13-OCT-2000	2000US -0239937
PR	20-OCT-2000	2000US -0240960
PR	20-OCT-2000	2000US -0241221
PR	20-OCT-2000	2000US -0241785
PR	20-OCT-2000	2000US -0241785
PR	20-OCT-2000	2000US -0241787
PR	20-OCT-2000	2000US -0241808
PR	20-OCT-2000	2000US -0241809
PR	20-OCT-2000	2000US -0241809
PR	20-OCT-2000	2000US -0241826
PR	01-NOV-2000	2000US -0246167
PR	08-NOV-2000	2000US -0246474
PR	08-NOV-2000	2000US -0246475
PR	08-NOV-2000	2000US -0246476
PR	08-NOV-2000	2000US -0246477
PR	08-NOV-2000	2000US -0246478
PR	08-NOV-2000	2000US -0246523
PR	08-NOV-2000	2000US -0246524
PR	08-NOV-2000	2000US -0246525
PR	08-NOV-2000	2000US -0246527
PR	08-NOV-2000	2000US -0246527
PR	08-NOV-2000	2000US -0246528
PR	08-NOV-2000	2000US -0246532
PR	08-NOV-2000	2000US -0246603
PR	08-NOV-2000	2000US -0246610
PR	08-NOV-2000	2000US -0246611
PR	08-NOV-2000	2000US -0246611

PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCT INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488783/53.  
XX P-PSDB; AAU16619.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
PI diagnosing, preventing, treating or ameliorating medical conditions and  
PI used as food additives or preservatives -  
XX  
XX  
PS Claim 1; SEQ ID No 785; 980bp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also, used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoabsorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. Rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease. Infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence encodes a novel secreted protein of the invention.

**Query Match**

9.68; Score 224; DB 22; Length 424;







XX 03-OCT-2001; 2001WO-US30821.  
 XX 03-OCT-2000; 2000US-237189P.  
 XX (GENE-) GENE LOGIC INC.  
 PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 PI MPI; 2002-435328/46.  
 XX  
 XX Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity -  
 PS Claim 1; SEQ ID NO 733; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicative of GCA.  
 CC Also included are modulating (M2) GCA by contacting GC with an agent  
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC modulating GCA; M3 is useful for screening an agent capable of modulating  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease; also bacterial infection, viral infection,  
 CC parasitic infection, protozoal infection, fungal infection and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1741 BP; 481 A; 482 C; 408 G; 370 T; 0 other;

XX Query Match 8.9%; Score 208.2; DB 24; Length 1741;

XX Best Local Similarity 53.9%; Pred. No. 1e-36; Mismatches 503; Indels 72; Gaps 9;

XX 301 GATTACTATTAAGATCAATTAACCCCTATGATATGGAACAAATTAAGACCGCTTGAA 360  
 XX 414 GATTATACCAAAATTAATAACACGCTATGACATGATGATTAAGAGAGACTGAA 473  
 XX 361 AACAACTTATGAGTATGCTCAGAGATGATCCAGACTTCAACACTATGATTAACAAT 420  
 XX 474 AACATTAATATGCTGCTGCTCAGAGTATGATCAAGATTTTAATACCATTTTACCAAC 533  
 XX 421 TGTTCATCTACCAACAGCCGTGAGATGACATAGCTTAAATGCGAAGAGCTTGGAAG 480  
 XX 534 TGTTCATCTACCAACAGCCGTGAGATGACATAGCTTAAATGCGAAGAGCTTGGAAG 593

QY 481 CTCTTCTTGCAAAAAATTAATGACCTACCAAGAGAAACCGAGATCATGATGCTCAG 540  
 Db 594 ATATTCCTACGAGAGGTTGCATCATGATGCAACAGAAAGAGATGATGATGATC 653  
 QY 541 GCAAAAG 600  
 Db 654 CTTAGAAACAGCCACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 713  
 QY 601 GTACCAAAACACATCAAGCTGATGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
 Db 714 GCCCATGAGTGGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 767  
 QY 661 CCTCTGTGAG 720  
 Db 768 CCACCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 799  
 QY 721 ACCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 Db 800 TCCCAACCCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859  
 QY 781 CAGCAGCAACCCCGAG 840  
 Db 860 CTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 919  
 QY 841 GCGGCAACCCCGAG 900  
 Db 920 G-----GAG 951  
 QY 901 ACCCCACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 Db 952 GCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011  
 QY 961 ACCAAGTGGGCGAG 1020  
 Db 1012 CCGCTTCCCTATGAG 1071  
 QY 1021 CCGCAGCTGAG 1080  
 Db 1072 CCGCAGCTGAG 1130  
 QY 1081 TGGCAG 1140  
 Db 1131 AATTGCAATGAG-ATTTTGAAGAGATGATGATGATGATGATGATGATGATGATGATG 1189  
 QY 1141 TTCTACAGAGCTGTGAG 1200  
 Db 1190 TTCTATTAACAG 1249  
 QY 1201 CACCCATGAG 1260  
 Db 1250 CCCCCATGAG 1309  
 QY 1261 CAGAGAGTGGTGTGAG 1320  
 Db 1310 CAGAGAGTGGTGTGAG 1367  
 QY 1321 GACCATGAGTGTGAG 1380  
 Db 1368 GATACAGATGTTGAG 1424  
 QY 1381 AAGATGCGAG 1440  
 Db 1425 AAGATGCGAG 1484  
 QY 1441 CCC---ACCAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 Db 1485 GCGTGGCAATGCTTCAAGATGCTTCAAGATGCTTCAAGATGCTTCAAGATGCTTCA 1544  
 QY 1498 GAGCATGAGATGCTTCAAGATGCTTCAAGATGCTTCAAGATGCTTCAAGATGCTTCA 1544  
 Db 1545 GAGCAAG 1591





```

Db 306 GCGTAGAGAGTGTGGCCCCATCAAGCCCCCAGCAAGAACTTGGCTGACTCTCAGCA 247
QY 1035 GCACCCAGACACAGAGAGAGCAAGAGTCTCGAGCAGCTCAAGTCTGACAGCGGCAAT 1094
Db 246 ACAACACACAGAGCTMTAAGAAAGAAAGCTTTCAGACAGTAAACATTCATATGSCAT 187
QY 1095 CCTCAAGGAGATGTTGGCCAAGAAAGCAGCGCGGCTACGCCCTGGCCCTTTCACAAGCCTGT 1154
Db 186 TTTGAAGGAGTTACTCTCTAAGAGCAAGCTGCTATGCTTGGCTTCTATTAACCAAT 127
QY 1155 GGAAGTGGAGCACTGGGCTTACACGACTGATGATCATCAAGCAGCCCATGAGCAT 1214
Db 126 GGATGCTTTCGCACTGGCCTGCATGACTACATGACATCATTAAGCAGCCCATGAGCT 67
QY 1215 GAGCACAATCAAGTCTAAGTGAAGGCGCGTGAATACCGTGAATGCTCAGAGTTTGGTGC 1274
Db 66 CAGCACTGTCAGCGGAAGATGAGAAACCGTGATTAACGGGATGACACAGAGTTTGGCTGC 7
QY 1275 TGAGCT 1280
Db 6 TGATGT 1

```

Search completed: March 1, 2003, 13:46:34  
 Job time : 506 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:56:39 : Search time 87 Seconds  
(without alignments)  
8209.772 Million cell updates/sec

Title: US-09-700-590A-101

Perfect score: 1 gcaaggagcactagcatgctc.....ngccttaaaagtlnccca 2329

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCtus.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159.2	6.8	450	US-09-328-111-516	Sequence 516, App
2	104.8	4.5	429	US-09-397-787-243	Sequence 243, App
3	84.6	3.6	7218	US-08-232-463-14	Sequence 14, Appl
4	64.8	2.8	2223	US-08-257-073-4	Sequence 4, Appl
5	64.8	2.8	8147	US-09-514-247A-9	Sequence 9, Appl
6	63.4	2.7	320	US-09-165-264-14	Sequence 14, Appl
7	63	2.7	320	US-09-165-264-7	Sequence 7, Appl
8	62.8	2.7	318	US-09-165-264-12	Sequence 12, Appl
9	62.6	2.7	320	US-09-165-264-11	Sequence 11, Appl
10	62.6	2.7	16442	US-08-781-891-208	Sequence 208, App
11	61.4	2.6	320	US-09-165-264-13	Sequence 13, Appl
12	60.6	2.6	319	US-09-165-264-8	Sequence 8, Appl
13	59.4	2.6	13987	US-08-804-227C-13	Sequence 13, Appl
14	59.4	2.6	44377	US-08-804-227C-7	Sequence 7, Appl
15	59.4	2.6	44377	US-08-804-198-1	Sequence 1, Appl
16	59	2.5	4403765	US-09-103-840A-2	Sequence 2, Appl
17	58.2	2.5	12980	US-08-811-566-5	Sequence 5, Appl
18	58.2	2.5	12980	US-09-034-756-5	Sequence 5, Appl
19	57.2	2.5	1798	US-09-797-906-1	Sequence 1, Appl
20	57	2.4	2447	US-09-014-969-14	Sequence 14, Appl
21	56.2	2.4	153731	US-09-128-155-16	Sequence 16, Appl
22	55.6	2.4	43795	US-08-742-185-101	Sequence 101, App
23	55	2.4	356	US-08-520-678A-22	Sequence 22, Appl
24	55	2.4	356	US-08-897-126-22	Sequence 22, Appl
25	55	2.4	9646	US-08-811-566-1	Sequence 1, Appl
26	55	2.4	9646	US-09-034-756-1	Sequence 1, Appl
27	54.6	2.3	240	US-08-628-417-6	Sequence 6, Appl

28	54.6	2.3	452	US-09-397-787-147	Sequence 147, App
29	54.2	2.3	7326	US-08-194-468-1	Sequence 1, Appl
30	54.2	2.3	7326	US-09-514-247A-7	Sequence 7, Appl
31	54.2	2.3	7344	US-08-961-739-1	Sequence 1, Appl
32	54.2	2.3	80246	US-09-078-294-4	Sequence 4, Appl
33	54.2	2.3	80595	US-09-078-294-3	Sequence 3, Appl
34	53.8	2.3	3275	US-09-370-838-151	Sequence 151, App
35	53.8	2.3	9589	US-07-925-695-1	Sequence 1, Appl
36	53.8	2.3	9589	US-07-925-695-2	Sequence 1, Appl
37	53.6	2.3	1051	US-09-245-041-10	Sequence 10, Appl
38	53	2.3	12001	US-08-458-568A-11	Sequence 11, Appl
39	52.8	2.3	1447	US-09-443-041A-27	Sequence 27, Appl
40	51.8	2.2	2674	US-09-817-180-1	Sequence 1, Appl
41	51.6	2.2	260	US-08-520-678A-29	Sequence 29, Appl
42	51.6	2.2	260	US-08-897-126-2	Sequence 2, Appl
43	51.6	2.2	980	US-09-171-209-8	Sequence 8, Appl
44	51.2	2.2	376	US-08-623-906A-18	Sequence 18, Appl
45	51.2	2.2	405	US-08-299-074A-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-328-111-516
: Sequence 516, Application US/09328111
: Patent No. 6262333
: GENERAL INFORMATION:
: APPLICANT: Endege, Wilson O.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon H.
: APPLICANT: Burgess, Christopher C.
: APPLICANT: Bushnell, Steven E.
: APPLICANT: Carroll III, Eddie
: APPLICANT: Catino, Theodore J.
: APPLICANT: Dertl, Adnan
: APPLICANT: Ford, Donna M.
: APPLICANT: Lewis, Marcia E.
: APPLICANT: Monahan, John E.
: APPLICANT: Schlegel, Robert
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: TITLE OF INVENTION: PRODUCTS
: FILE REFERENCE: CCD-257 (US)
: CURRENT APPLICATION NUMBER: US/09/328,111
: CURRENT FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: US 60/088,801
: EARLIER FILING DATE: 1998-06-10
: NUMBER OF SEQ ID NOS: 850
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 516
: LENGTH: 450
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)..(450)
: OTHER INFORMATION: n = A,T,C or G
US-09-328-111-516

Query Match 6.8%, Score 159.2; DB 4; Length 450;
Best Local Similarity 69.8%; Pred. No. 4,1e-31;
Matches 215; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 975 GCGGGGAGAGACACCGCCCTGTGAACCTCCAAAGAGACGTCGCCGACTCTACACA 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 GGTAGAGAGAGGTGTGCCCCCATCAAGCCCCGCAAGAGCTTGCTGACTCTACACA 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1035 GCACCCAGCACACAGAGAGACGACGACGAGCTCAAGTGTCTGACGCGCAT 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 ACAACACGAGAGCTCTAAGAAAGAAAGCTTTACAGAAAGTAAACATTGCAATGGCAT 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1095 CCTCAGAGAGATGTTGGCAAGAACGACGCGCTTACGCTTGCGCCCTTCTACAAACCTGT 1154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

DB 242 TTGGAAGAGTACTCTCTAAGACATGCTGCTATGCTTGGCCTTTCTATAACAGT 301  
QY 1155 GGACGTGGAGGACATGGCCCTACACACTACTGTGACATCATCAAGCCCATGACAT 1214  
DB 302 GATGCTTCTGCACTTGGCTGCTGACTACATCATCATTAAGACCCCATGGACCT 361  
QY 1215 GACACATTAATCTAATCTGAAGCCCTGACTACCTGATGCTCAGAGATTGTGTC 1274  
DB 362 CACACATGTCACCGGAAGATGAGAACCGTGTATTCACGGGATGCACAGAGTTTGTGC 421  
QY 1275 TGAGTCC 1282  
DB 422 TGATGTAC 429

RESULT 2  
US-09-397-787-243  
Sequence 243, Application US/09397787  
Patent No. 6468758  
GENERAL INFORMATION:  
APPLICANT: Benson, Darin R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN  
TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS  
FILE REFERENCE: 210121.466C2  
CURRENT APPLICATION NUMBER: US/09/397,787  
CURRENT FILING DATE: 1999-09-16  
NUMBER OF SEQ ID NOS: 334  
SOFTWARE: FastSeq for windows Version 3.0  
SEQ ID NO 243  
LENGTH: 429  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-397-787-243

Query Match 4.5%; Score 104.8; DB 4; Length 429;  
Best Local Similarity 72.3%; Pred. No. 2.6e-17;  
Matches 136; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1810 TATGAGTCGAGGAGAGGACAGTCAAGCTATGCTATGAGGAGAGCGGACCTC 1869  
DB 242 TATGATTCGAGGAGAGGAGAGAGAGAGAGCCATGATGATATGAGAGACAGCTTA 301  
QY 1870 AGCTTGACATCAACAAGCTCCCGGAGAGAGCTGGCGCGCTGTGTCATCATCCAG 1929  
DB 302 AGCTGATATCAATTAAGTACTCTGAGGAGAAAGCTGGGTCGATGATATCATCCAA 361  
QY 1930 TCACGGGAGCCCTCCCTGGAAGATTCCACCCCGGAGAGATTTGAATCCACTTTGAGACC 1989  
DB 362 GCCAGGGAACCTCTCTAGCTGATCAAAATCCAGAAATGAGATTGATTTTGAAMCA 421  
QY 1990 CTGAGCC 1997  
DB 422 CTCAGCC 429

RESULT 3  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA

COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PT9pbt-F15  
US-08-232-463-14

Query Match 3.6%; Score 84.6; DB 1; Length 7218;  
Best Local Similarity 3.1%; Pred. NO. 1.7e-11;  
Matches 12; Conservative 247; Mismatches 126; Indels 0; Gaps 0;

QY 1508 GTTCAGCTGATGATCTGAGGAGAGAGCCCGCGGTGCTGAGCTCCAGAGCAGC 1567  
DB 1438 GTACRR 1379  
QY 1568 TCAAGCCGTCGACGAGAGCTTGACGCTCTCTAGCCCCGACGAAACAACAAGA 1627  
DB 1378 RRR 1319  
QY 1628 AAAGGAGAAAGCAAGAGAAAGAAAGAAAGCAAGAAAGAGAGAGAGTG 1687  
DB 1318 RRR 1259  
QY 1688 AAGAGATATAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1747  
DB 1258 RRR 1199  
QY 1748 GCAGCAACCAATGTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1807  
DB 1198 RRR 1139  
QY 1808 CGATGAGTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1867  
DB 1138 RRR 1079  
QY 1868 TCAGCTTGACATCAACAAGCTCC 1892  
DB 1078 RRRRRRRRRRRRRATGCAAGCTCC 1054

RESULT 4  
US-08-257-073-4  
Sequence 4, Application US/08257073  
Patent No. 576597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo





QY 773 TGGCCCCGAGCAGCAACCCCGCTCCAGCTCCCGGATGACAGACCCAC 832  
Db 140 CC 81  
QY 833 CCATCATCGGCGCCACCCAGAGCC 857  
Db 80 CC 56

## RESULT 7

US-09-165-264-7/c

; Sequence 7, Application US/09165264

; Patent No. 6197510

; GENERAL INFORMATION:

; APPLICANT: Vinayagamorthy, Thuraiayah

; TITLE OF INVENTION: Multi-Loci Genomic Analysis

; FILE REFERENCE: 44747

; CURRENT APPLICATION NUMBER: US/09/165,264

; CURRENT FILING DATE: 1998-10-01

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent Ver. 2.1

; SEQ ID NO 7

; LENGTH: 320

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence  
US-09-165-264-7Query Match 2.7%; Score 63; DB 4; Length 320;  
Best Local Similarity 52.9%; Pred. No. 9,3e-07;

Matches 135; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 603 ACCAACAACAATCAATGATGCTCTCTCGAGACCCGATCGGATCTCC 662  
Db 320 ACTAAGGAGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 261  
QY 663 TCTGTGAGGCGGCTCAACCTTCTCTGCTGCTGCTGCTGCTGCTGCT 722  
Db 260 CC 201  
QY 723 CCTGTGATGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 782  
Db 200 CC 141  
QY 783 GCGACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842  
Db 140 CC 81  
QY 843 GCGACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857  
Db 80 CC 66

## RESULT 8

US-09-165-264-12/c

; Sequence 12, Application US/09165264

; Patent No. 6197510

; GENERAL INFORMATION:

; APPLICANT: Vinayagamorthy, Thuraiayah

; TITLE OF INVENTION: Multi-Loci Genomic Analysis

; FILE REFERENCE: 44747

; CURRENT APPLICATION NUMBER: US/09/165,264

; CURRENT FILING DATE: 1998-10-01

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent Ver. 2.1

; SEQ ID NO 12

; LENGTH: 318

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence  
US-09-165-264-12Query Match 2.7%; Score 62.8; DB 4; Length 318;  
Best Local Similarity 53.2%; Pred. No. 1e-06;  
Matches 133; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 608 ACACAACTGAGATGATCTCTCTCGAGACCCGATGACGCGAATCTCTCT 667  
Db 311 ACACAACTGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 252  
QY 668 TGCAGGCGACGCTCAACCTTCTCTGCTGCTGCTGCTGCTGCTGCTG 727  
Db 251 CC 192  
QY 728 TCAATGAGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787  
Db 191 CC 132  
QY 788 AACCCCGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847  
Db 131 CC 72  
QY 848 CCGCAGAGCC 857  
Db 71 CCCCCCCCCC 62

## RESULT 9

US-09-165-264-11/c

; Sequence 11, Application US/09165264

; Patent No. 6197510

; GENERAL INFORMATION:

; APPLICANT: Vinayagamorthy, Thuraiayah

; TITLE OF INVENTION: Multi-Loci Genomic Analysis

; FILE REFERENCE: 44747

; CURRENT APPLICATION NUMBER: US/09/165,264

; CURRENT FILING DATE: 1998-10-01

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent Ver. 2.1

; SEQ ID NO 11

; LENGTH: 320

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence  
US-09-165-264-11Query Match 2.7%; Score 62.6; DB 4; Length 320;  
Best Local Similarity 51.6%; Pred. No. 1,2e-06;  
Matches 143; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 581 CAAACCTGAGCTTCTCAAGGATGCAACCACTGACATGATGCTCTCGAGAGCC 640  
Db 310 CAAATCTGCT 251  
QY 641 AGACCGCTCAGCGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 700  
Db 250 CC 191  
QY 701 CCGGAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760  
Db 190 CC 131  
QY 761 CCGGCGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 820  
Db 130 CC 71  
QY 821 AGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 857  
Db 70 CC 34

## RESULT 10

US-08-781-891-208/c





OY	654	GAATCTCTCTCTGTGGAGGCACAGGCTCAACCCCTTCCCTGGCGGACCCCGGAGCTCAT	713
Db	9948	AACCACGCCCCCATACCAACCCACTGGCGACCTTGACCCGTTAAACAACCCCCAACCCCA	9889
OY	714	CGTCCAGACCCCTGTCTCATGTAGAGTGGTGGCTTCCCGACGCATGTACAGAGCCGCCCGCAGT	773
Db	9888	CGACCCCGGCACACGGCACTCTCCCGGACACACAGCACGGGCCCCCGGAAGACACACCCACAGA	98239
OY	774	GCCCCCGGCAGCACAACCCCGCACCGCTCGAGTCTCCCAAGCCCGTACAGAGCCACCCACC	833
Db	9828	GCCCCGGAGAGCACAACATCATCAAGACCCGGCAGAACTCTCAAGGCCCGCGGCCACACAC	9765
OY	834	CATCATGTGGCGGCCACCCCAACGCTTGTGAAGACAAGAAAGGAGTGTGAAGAGGAAGAAACGA	893
Db	9768	ACCCGACGCTTACCGAAGCAGCAGCAGCAGCACACACCAACAGACCCCGCCCATCTCCGCGGA	9709
OY	894	CACGACACCCCAACCAACATTGACCCCAATTACAGAGCCACCCCTGGCTGCCCGCGAGACC	953
Db	9708	TCTCCACCCCAACCGGCTTCACACACTCAGCAACACGGCACCCTGGCGCGCAACGCCGCC	9649
OY	954	CAAGACACCAACG 966	
Db	9648	TTCGACCGCGCAG 9636	

### RESULT 15

```

1 Sequence 1, Application US/08804198
2 Patent No. 5945320
3 GENERAL INFORMATION:
4 APPLICANT: Burgelt, Stanley G.
5 APPLICANT: Appleton, Stanley G.
6 APPLICANT: Appleton, Stanley G.
7 APPLICANT: Rao, Madanraj A.
8 APPLICANT: Richardson, Mark A.
9 APPLICANT: Rostek, Paul R., Jr.
10 TITLE OF INVENTION: PLATENOLEIDE SYNTHASE GENES
11 NUMBER OF SEQUENCES: 6
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: PAUL R. CAMTRELL 1138
14 STREET: LILLY CORPORATE CENTER
15 CITY: INDIANAPOLIS
16 STATE: IN
17 COUNTRY: USA
18 ZIP: 46285
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: Macintosh
22 OPERATING SYSTEM: Macintosh 7.0
23 SOFTWARE: Microsoft word 5.1
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/804,198
26 FILING DATE:
27 CLASSIFICATION: 435
28 ATTORNEY/AGENT INFORMATION:
29 NAME: CAMTRELL, PAUL R.
30 REGISTRATION NUMBER: 36,470
31 REFERENCE/DOCKET INFORMATION: P9113
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 317-276-3885
34 INFORMATION FOR SEQ ID NO: 1:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 44377 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (genomic)
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 350..14002
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 14046..20036
47 FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 20110..31284
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
;
US-08-804-198-1

```

Query Match	Score	DB 2;	Length
2.68;	59.4;	DB 2;	44377;

Matches 177; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY	594	TTTCAGGTTGAAAACACACTCAACATCGACTTCCTCGAGACCACGACGACCCCTACGC	653
Db	10008	TCGCCCAACACACCGACAAACACTTATCTACGACTCTCGGAAAACCCCACTCCCGCATAC	9949
QY	654	GAATCTCTCTCTGTGTGACGACAGGCTCAACCCCTTCCCTGCCGTACACCCTGGAACTCAT	713
.Db	9948	AACCCAGGCCCATATACCAACCCACTGTGGGACCTGTAGCCGTGAACACACCCCAACCCCA	9889
.QY.	714	CGTCAACACCCCTGTCTATGTACAGTGGTGCCTCCCAAGCCACTGTCAAGAGCCCCGGCAGT	773
Db	9888	CGACCCCGGACACCGCACTCCCCCGCCACACAGACCGGGCCCCGGAAGACACACCCACAGA	9829
QY	774	GGCCCCGACGACAAACCCCCACCGGCTTCAGTCTCCCAAGCCGTACAGACGACCCACCC	833
Db	9828	GGCCCCGAGGACACACATATCAAGACCCGGCAGCACTCTCAAGGCCCCCGGCAACCCAC	9769
QY	834	CATCATGTGCGGGACACCCCAACAGCCTTGTGAAGACAAAGAGGAGTGAAGAGAAAGCAGA	893
Db	9768	ACGGCAAGCTTCACGAAAGACAGACGACGACACCAACCAAGACCCCGCCACATCCCGGGA	9709
QY	894	CACCAACACCCCAACACCACTTGAAGCCCATTCACGAGCCACCCCTGTGCTGCCCGGAGCC	953
Db	9708	TCCGCACCCCAACCGGTCCACACACTCAGCAACACGGACCGCTGGCGCCGCAAGCCGCC	9649
QY	954	CAGACCAACCAAG	966
Db	9648	TCGACACGCGCAG	9636

Search completed: March 1, 2003, 14:44:56  
Job time : 244 secs

Job time : 244 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:57:34 ; Search time 140 Seconds  
(without alignments)  
10374.344 Million cell updates/sec

Title: US-09-700-590A-101  
Perfect score: 2329  
Sequence: 1 gcaggatcactagcatgctc.....ngccttaaaaggttccccca 2329

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 460893 seqs, 311809382 residues

Total number of hits satisfying chosen parameters: 921786

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*  
1: /cgn2\_6/ptoddata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptoddata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptoddata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptoddata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptoddata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptoddata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptoddata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptoddata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptoddata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptoddata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptoddata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptoddata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptoddata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptoddata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	460.2	19.8	4664	10	US-09-880-107-1657 Sequence 1657, App
2	224	9.6	424	10	US-09-764-864-785 Sequence 785, App
3	224	9.6	424	10	US-09-764-864-372 Sequence 372, App
4	159.2	6.8	450	10	US-09-879-536-516 Sequence 516, App
5	137	5.9	358	10	US-09-783-590-1966 Sequence 1966, App
6	104.8	4.5	429	10	US-09-876-889-243 Sequence 243, App
7	66.6	2.9	244	10	US-09-925-301-299 Sequence 299, App
8	64.8	2.8	814	12	US-10-109-886-9 Sequence 9, App
9	64	2.7	463	10	US-09-960-352-7186 Sequence 7186, App
10	63.6	2.7	681	10	US-09-822-830A-26 Sequence 26, App
11	63.6	2.7	12733	9	US-10-032-393-47 Sequence 47, App
12	63.6	2.7	12739	9	US-10-032-393-8 Sequence 8, App
13	63.4	2.7	8895	10	US-09-764-853-887 Sequence 887, App
14	63.4	2.7	8895	10	US-09-764-853-937 Sequence 937, App
15	63.4	2.7	9656	10	US-09-764-853-886 Sequence 886, App
16	63.4	2.7	9656	10	US-09-764-853-933 Sequence 933, App
17	62	2.7	423	10	US-09-864-761-18355 Sequence 18355, A
18	61.4	2.6	283	10	US-09-960-352-9095 Sequence 9095, App
19	61	2.6	217	10	US-09-960-352-6539 Sequence 6539, App

20	61	2.6	2269	9	US-09-989-920-113 Sequence 113, App
C 21	60.8	2.6	341	10	US-09-960-352-12302 Sequence 12302, A
C 22	60.6	2.6	446	10	US-09-960-352-3400 Sequence 3400, App
C 23	60.2	2.6	415	10	US-09-960-352-2223 Sequence 2223, App
C 24	60.2	2.6	2608	10	US-09-739-254-16 Sequence 16, App
C 25	60.2	2.6	2608	10	US-09-904-615-16 Sequence 16, App
C 26	59.4	2.6	378	10	US-09-960-352-14879 Sequence 14879, A
C 27	59.4	2.6	475	10	US-09-864-761-6203 Sequence 6203, App
C 28	59.4	2.6	511	10	US-09-864-761-22817 Sequence 22817, A
C 29	59.4	2.6	2903	9	US-09-822-846-239 Sequence 239, App
C 30	59.2	2.5	547	10	US-09-815-343-213 Sequence 213, App
C 31	58.6	2.5	305	10	US-09-864-761-19262 Sequence 19262, A
C 32	58.6	2.5	496	10	US-09-864-761-2534 Sequence 2534, App
C 33	58.4	2.5	261	10	US-09-864-761-19524 Sequence 19524, A
C 34	58.4	2.5	261	10	US-09-864-761-25517 Sequence 25517, A
C 35	58.4	2.5	448	10	US-09-864-761-557 Sequence 557, App
C 36	58.4	2.5	451	10	US-09-860-352-10262 Sequence 10262, A
C 37	58.4	2.5	513	10	US-09-864-761-8851 Sequence 8851, App
C 38	58.4	2.5	4237	10	US-09-745-763-20 Sequence 20, App
C 39	58.2	2.5	276	10	US-09-864-761-20595 Sequence 20595, App
C 40	58.2	2.5	529	10	US-09-983-965-2109 Sequence 2109, App
C 41	58.2	2.5	12980	9	US-09-985-937-5 Sequence 5, App
C 42	58.2	2.5	12980	10	US-09-238-076-5 Sequence 5, App
C 43	57.6	2.5	385	10	US-09-960-352-6420 Sequence 6420, App
C 44	57.6	2.5	425	10	US-09-960-352-4010 Sequence 4010, App
C 45	57.6	2.5	456	10	US-09-960-352-9659 Sequence 9659, App

## ALIGNMENTS

RESULT 1  
US-09-880-107-1657  
Sequence 1657, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OR INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-MO  
CURRENT APPLICATION NUMBER: US/09/880,107  
PRIOR FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1657  
LENGTH: 4664  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 D42040  
US-09-880-107-1657  
Query Match 19.8%; Score 460.2; DB 10; Length 4664;  
Best Local Similarity 57.0%; Pred. No. 1.1e-101;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;  
QY 143 CCAACCCCCCCCCCGAGAGCCTCAACCTCAACGAGCCCAAGAGGAGAGCAACCAAC 202  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 1877 CCAACCCACACACCCCGAGAGTGTTCATTCCTCAACCAAGGAGAGGAGTTCACCAACGAGC 1936  
QY 203 TGCATATCCTGCTCAGAGTGTCTCAAGACATATGAAACACACAGTTGGACGCTT 262  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 1937 TGCATATCCTGCTCAGAGTGTCTCAAGAGGCTGTGGAACATCAGTTCGATGCGCAT 1996  
QY 263 TCCAGCAGCCTGTGATGCGCTCAAGCTGAACCTCCCTGATTAATTAAGTCTTAA 322  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 1997 TCCGAGAGCCTGTGATGCTGTCAAACTGGCTTACCGAGATTATACCAAAATTTATAAAC 2056

```
QY 323 CCGCTATGATATGGACAAATTAAGACGCTTGGAAAAACAATTAAGTACTGGAATGCTC 382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2057 ACCCTATGACATGGGTACTATTAAAGAGAGACTTGAACAAATTTATTTATGGCTGCTT 2116
QY 383 AGGAATGATCCAGAGCTTCAACACTATGTTTCAAAATTTGTTACATCTACAAAGCTTG 442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2117 CAGAGGTATGCAAGATTTTAAATACATGTTCCCAACTGTATCAATTTTCAACAAAGCCCA 2176
QY 443 GAGATGACATAGTCTTAATGCGAAGCTCTGGAAAAAGCTCTTTCGAAAAATTAATG 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2177 CTGATATATTTGCTTAATGCGCAAAAGCGTGGAAAAAGATTTCTTACAGAAAGTTGCA 2236
QY 503 AGCTACCCACAGAAACCGAGATCATGATAGTCCAGGCAAAAAGAGAGAGAGAGAGAG 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2237 CATGCGACAGAGAAAGAAAGAGAGCTGTAGTGACCATCCCTTAAGAACAGCCACAAAGAG 2296
QY 563 GGAAGAAACAGGAGACAGCAAAACCTGGGGTTTCCAGGGTACCAACAAACACAGCAAT 622
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2297 GGGCCAAAGTTGGCAGCGCTCCAGGGCAGTGTACAGTGCCCTCAGTGCCCTGCGCTCT 2356
QY 623 CGACTCTCCGAGACACCCAGACCCCTCAGCCGAACTCTCTCTCTGTCAGGCGCCGCTC 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2357 -----CTTCTGTGTACACAGACCCCTGTATATCTCTCCAGCTGAGATACCTACCACTG 2410
QY 683 ACCCTTCCCTGCGCTGACCCCGGAGCTCATGCTCCAGACCCCTGTATGACAGTGGTGC 742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2411 TCCCTCAACATTCCTCCACCCCATCATGCTATTTCC-----T 2443
QY 743 CTCCTCCAGCCTGAGAGACCCCGCCCAAGTCCCGCCCAAGCAACCCCGCCGCTC 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2444 CTCCTATTTCTCAAGTCTTGTGACATGCTGTGTGAGACCCCGCTCTGCTGTATGTCAGCTC 2503
QY 803 CAGCTTCCAGCCCGCTGACAGAGCCACCCATCATGCGGGCCACCCAGACCTGTGTA 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2504 CTCACAGCCAGCCCTTGTGCGAAGAAAAAGGCGT----- 2537
QY 863 AGACAAAGAAAGGAGTGAAGAGAAAGAGAGACACCCACCCACCCACCATTTGACCCCA 922
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2538 ----AAAGGGAGAAAGAGATTAACACCCCTACACCTACAGCCTTGGCTCTGCTG 2593
QY 923 TTCACAGACACCCCTGCTGCGCCCGGAGCCCAAGACCAACCAAGCTTGGGCG----AGCGG 979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2594 CTCACAGCTAGCCCTCTCTGGGAGCTTGAAGCTTAAGGACAGAGGCTTCCCTCATGCTGTA 2653
QY 980 GGGAGAGACAGCGGCTGTGAAACCTCAAAAGAGAGAGCTGCCGACTCTACAGACACCC 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2654 GAGAGAGTGGTGGCCCTCAAGCCCGCCAGCAAGAGACTTGTGACTCTACAGCAACAC 2713
QY 1040 CAGCAACCAAGAAAGACACAGAGTCTCGGAGAGAGCTCAAGTGTCTGACGCGGATCTCTCA 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2714 ACCAGAGCTTAAGAAAGAAAGCTTTCAGAAAGTTAAACATTTCAATGCGATTTTGA 2773
QY 1100 AGGAGATGTTTGGCAAGAGACGCGGCTAGCGCTTGGCCCTTCTCAAGCTGTGAGAG 1159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2774 AGGAGTACTCTCTAAGAAAGCAATGCTGCTATGCTTGGCTTCTTAAACAGTGGAGT 2833
QY 1160 TGGAGGCACTGGGCTTACACAGCTAGTGTACATCAATCAAGCAACCCCATGAGCATAGCA 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2834 CTTCGACACTTGGCCGTGATGACTACATGACATCAATTAAGACCCCATGAGACTCAGCA 2893
QY 1220 CAAATCAAGCTAACTGAGAGCCCGGAGTACCGTATACGCTGAGAGTTTGGTGTGAGG 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2894 CTGTCAAGGGGAGAGATGAGAAACCGTATTAACCGGATGACACAGGAGTTGCTGTGAGT 2953
QY 1280 TCCGATGATGTTTCTCACTGCTATTAATACAAACCTCTGACATGAGTGGTGGCCA 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2954 TACGGCTTATGTTCTCAACTGCTATTAATGATACATCCCAAGTACGATGTTGTGGCAA 3013
QY 1340 TGGCCCGCAAGCTCCAGAGTGTGTTGAAATGCGCTTTCAGAGATCCGCGAGAGAGCTG 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3014 TGGCAAGAAAGCTACAGAGTATTTGAGTTCGGTATGCGCAAGATGCCAGATGAACACAC 3073
```

```
QY 1400 AGGAGCCAGTGTGCGCTGTCTCCCGGCACTGCCCCCTCC--ACCAAGTTGTGG 1456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3074 TGAACCAAGGGCCCTTACCAGTCTTACTGCGCATGGCCCTGGCCCAATGCTTT 3133
QY 1457 CCCCCGCTTCATCCAGCAGCAGCAGCAGGATAGCTCTC----- 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3134 CAGAGTCTTCAGTGAAGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3193
QY 1497 -----GGACAGTACAGTTCGACTGATACCTGAGTGTGAGAGAGAGAGAGAGAGAG 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3194 ATGAGAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3253
QY 1541 AGCGCTGCTGAGTCTCCAGAGACAGCTCAAAAGCCGTGACAGACAGCTTGCAGCCCTCT 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3254 ATGCTTACAGCAAGATCAAGAGAAACAGCTTGGCGCAGTACATGAACTGCGCTGTGT 3313
QY 1601 CTCAGCCCGCAGCAGCAACCAACCAAGA-AAAAGCAAAAGACAGAGAGAGAGAGAGAG 1659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3314 CCGAGGGTCCATATATCCAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3373
QY 1660 GAAAGCACAAGAAAGAGAGAG-----GAAGTGAAGAGATTAAGAGAGAGAGAGAGAG 1705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3374 GGAAGGCAAGAAAGATCGAGGCGGAGCTGGGGCGATGAAATGACAAAGGGCCCTAGGG 3433
QY 1706 AAGCCAGAGACCTCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3434 CACCCCGCCACCTCAACCTAAGAAAGTCCAAAGAAAGTGGCAGTGGGGGTGCGCAGT 3493
QY 1750 -----AGCAACAGCATGTGACGACA 1771
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3494 CTGCTTAAAGCCCTTGTGCTTGTGACCTTCTGAGAGAGAGTGGACCAAGCTCCCAAAA 3553
QY 1772 AGGAGCCAGCGCCCATGAAGAGAGCAAGCAAGCCCTCCACGATGAGTGTGAGAGAGAG 1831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3554 AGGCCACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3613
QY 1832 AGTGCAGACCTATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3614 AGAGCAGGGCCATGAGTTCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3673
QY 1892 CCGGCGAGAGAGTGGGCGCGGTGTCACATCATCCAGTCAAGGAGAGAGAGAGAGAGAG 1951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3674 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3733
QY 1952 ATTCCAGCCCGCAGATTTGAATGACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3734 ATTCAAGCCAGAGAGAGATTTGAGATTTGAATTTGAACACTCAAGCCATCCACACTTAG 3793
QY 2012 AGCTTGGAGCGCTATGCTACCTCTGTTGGCGAAGAAA 2050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3794 AGCTT-GAGCGCTATGCTCTTCTGCTTACGTAAGAAA 3831
```

## RESULT 2

US-09-764-864-785

Sequence 785, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic acids, proteins, and antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 785

LENGTH: 424

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-864-785

Query Match

9.6%; Score 224; DB 10; Length 424;





```
Db 242 TTTGAAGGACTTACTCTCTAAGAGCATGCTGCTTATGCTTGGCTTTCTATTAACCAT 301
OY 1155 GGACGTGGAGGACACGTGGCCCTACACGATCTGACATCATCAAGCACCACATGACAT 1214
Db 302 GGATGCTTCTGACATGCTGCTGATGATCATCATGACATCATTAAGACACCACATGACCT 361
OY 1215 GAGCAATATCAATCTTAATGAGAGCCCTGATGATGATGATGATGATGATGATGATGAT 1274
Db 362 CAGCACTGTCACGGAAGATGAGAACCTGATTAACCGGATGACACAGAGTTTGTCTGC 421
OY 1275 TGACGTCC 1282
Db 422 TGATGTAC 429

RESULT 5
US-09-783-590-1966
; Sequence 1966, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16,2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1966
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (102)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (121)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (133)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (158)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (180)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (186)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (203)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (225)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (237)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (331)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (334)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (336)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (339)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (347)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (355)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1966
```

Query Match 5.9%; Score 137; DB 10; Length 358;  
Best Local Similarity 82.3%; Pred. No. 8,3e-24;  
Matches 237; Conservative 0; Mismatches 39; Indels 12; Gaps 7;

```
OY 881 AGAGAAAGACAGACACACACACACACACACACACACACACACACACACACACACACACAC 940
Db 6 AGGNNAAAGACAGACACACACACACACACACACACACACACACACACACACACACACACAC 65
OY 941 TGCCCCGGAGCCCAAGACACCAAGCTGGCCGAGCGGAGAGACAGCCGCTGTG- 999
Db 66 TGCCCCGGAGCCCAAGACACCAAGCTGGCCGAGCGGAGAGACAGCCGCTGTG- 125
OY 1000 --AAACCTCAAGAAGAGAGCTGCCGACTCT-CAGCAGCACCACACACAGAA--GA 1054
Db 126 GAAACCTNCCAAAGAGAGAGCTGCCGACTCTNACACAGACACACACAGAGNAGAG- 185
OY 1055 GCAGCAAGGCTCGAGAGC--AGCTCAAGTGTCTCAGCGG-CATCTCAAGGA-GATGTTT 1110
Db 186 NCAGCAAGGCTCGAGAGCAGCTTCAAGTGTCTCAGCGGNCATCTTCAAGGAGATGTTT 245
OY 1111 GCCAAGACGCGCGCTTACGCT--GGCCCTTCTACAAGCCTGTG- 1156
Db 246 GCCAAGACGCGCTTACGCTTGGCCCTTCTACAAGCCTGTG- 293
```

```
RESULT 6
US-09-876-889-243
; Sequence 243, Application US/09876889
; Patent No. US2002007615A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; FILE REFERENCE: 210121.466C3
; CURRENT APPLICATION NUMBER: US/09/876,889
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-876-889-243
```

Query Match 4.5%; Score 104.8; DB 10; Length 429;  
Best Local Similarity 72.3%; Pred. No. 6e-16;





```

QY 622 TCAGCTCTCCGCAAGACCCCTCAGCCGAATCTCTCTGTGACAGCCAGCCT 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5358 TCGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5299
QY 682 CACCCCTTCCCGCCGTCACCCCGGACCTCATGTCGACAGACCCCTGTATGACGTGTG 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5298 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5239
QY 742 CCTCCCGACGACCTGACAGACGCCCGGACAGTGGCCCCCGACCAACCCCGACCCGCT 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5238 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5179
QY 802 CCACTCCCGACGCGGACAGACGACCCACCATCATGCGGACCCACCAAGCCTGTG 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5178 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5119
QY 862 AAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5118 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5059
QY 922 ATTCAGGAGCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 963
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5058 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5017
```

## RESULT 13

```

; US-09-764-853-887/c
; Sequence 887, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 887
; LENGTH: 8895
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-853-887
```

```

Query Match      2.7%; Score 63.4; DB 10; Length 8895;
Best Local Similarity 52.9%; Pred. No. 3.3e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
```

```

QY 1610 AGCAGAACCAACCAAGAAAAAGAGAAAGCAAGAGAAAAAGAAAAAGAACACA 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 945
QY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
QY 1730 AGACGAAGAAAAATATATAGCAGCAACATGTGACGAAGAGAGAGAGAGAGAGAGAG 1789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 825
QY 1790 AGACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 1850 ATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 AGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 748
```

```

RESULT 14
US-09-764-853-937/c
; Sequence 937, Application US/09764853
```

```

; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 937
; LENGTH: 8895
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-853-937
```

```

Query Match      2.7%; Score 63.4; DB 10; Length 8895;
Best Local Similarity 52.9%; Pred. No. 3.3e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 1610 AGCAGAACCAACCAAGAAAAAGAGAAAGCAAGAGAAAAAGAAAAAGAACACA 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945
QY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
QY 1730 AGACGAAGAAAAATATATAGCAGCAACATGTGACGAAGAGAGAGAGAGAGAGAGAG 1789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
QY 1790 AGACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 1850 ATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 AGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 748
```

## RESULT 15

```

; US-09-764-853-886/c
; Sequence 886, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 886
; LENGTH: 9656
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-853-886
```

```

Query Match      2.7%; Score 63.4; DB 10; Length 9656;
Best Local Similarity 52.9%; Pred. No. 3.4e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
```

```

QY 1610 AGCAGAACCAACCAAGAAAAAGAGAAAGCAAGAGAAAAAGAAAAAGAACACA 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1761 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
QY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1701 AGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1642
QY 1730 AGACGAAGAAAAATATATAGCAGCAACATGTGACGAAGAGAGAGAGAGAGAGAGAG 1789
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:52:00 ; Search time 3254 Seconds  
(without alignments)  
11591.668 Million cell updates/sec

Title: us-09-700-590a-101

Perfect score: 2329  
1 gcaggagcactacatgc.....ngccttaaaaggtinccca 2329

Scoring table: IDENTITY\_MNC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hnc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hnc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pla:\*  
21: em\_gss\_vtl:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.8	71.1	1716	11	BC000156 Homo sapi
2	1640	70.4	1743	11	BC030158 Homo sapi
3	1411	60.6	1776	11	BC011541 Mus muscu
4	856	36.8	987	14	BO651563 AGENCOURT
5	787	33.8	951	14	BO646784 AGENCOURT
6	779	33.4	893	14	BQ219777 AGENCOURT

7	770.2	33.1	933	14	BQ214403 AGENCOURT
8	756.6	32.5	975	14	BQ653187 AGENCOURT
9	752.4	32.3	925	14	BQ645818 AGENCOURT
10	736.2	31.6	899	14	BQ647311 AGENCOURT
11	735.8	31.6	931	14	BO650713 AGENCOURT
12	733.4	31.5	919	14	BQ652326 AGENCOURT
13	730.4	31.4	926	14	BQ650970 AGENCOURT
14	729	31.3	740	14	BM983041 AGENCOURT
15	715.2	30.7	938	14	BO654353 AGENCOURT
16	687.8	29.5	1106	13	BM452148 AGENCOURT
17	665.4	28.6	2085	11	BC010699 AGENCOURT
18	664.2	28.5	971	14	BQ652321 AGENCOURT
19	647.8	27.8	815	14	BQ749323 AGENCOURT
20	630	27.1	1055	12	BG824322 AGENCOURT
21	619	26.6	641	13	B1907231 AGENCOURT
22	613.8	26.4	627	14	BM679991 AGENCOURT
23	611.6	26.3	902	14	BQ232371 AGENCOURT
24	606.8	26.1	666	12	BF718610 AGENCOURT
25	606	26.0	709	13	B1414784 AGENCOURT
26	598.6	25.7	1011	11	BC013666 AGENCOURT
27	598.4	25.7	619	14	BQ939643 AGENCOURT
28	594.8	25.5	682	13	B1252631 AGENCOURT
29	592	25.4	615	12	BG391946 AGENCOURT
30	573	24.6	583	14	BQ674079 AGENCOURT
31	565.6	24.3	572	9	A1394011 AGENCOURT
32	552	23.7	645	10	BE294756 AGENCOURT
33	549.4	23.6	937	12	BG032535 AGENCOURT
34	547.6	23.5	675	12	BF550759 AGENCOURT
35	547	23.5	950	12	BF530953 AGENCOURT
36	544.8	23.4	548	9	A1480035 AGENCOURT
37	542.6	23.3	658	13	BM389207 AGENCOURT
38	536.4	23.0	631	12	BF712714 AGENCOURT
39	533.2	22.9	786	13	B1852791 AGENCOURT
40	529.2	22.7	919	12	BF180998 AGENCOURT
41	527.8	22.7	544	9	AL044273 AGENCOURT
42	527.8	22.7	944	14	BQ049332 AGENCOURT
43	527	22.6	535	9	A1360247 AGENCOURT
44	521.4	22.4	647	12	BF237405 AGENCOURT
45	520.6	22.4	1641	11	BC004832 AGENCOURT

#### ALIGNMENTS

RESULT 1  
BC000156  
LOCUS  
DEFINITION  
Homo sapiens, Similar to bromodomain-containing 4, clone  
IMAGE:2900081, mRNA.  
ACCESSION  
BC000156  
VERSION  
BC000156.1 GI:14704448  
KEYWORDS  
Homo sapiens.  
SOURCE  
Homo sapiens.  
ORGANISM  
Homo sapiens.  
REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1716)  
AUTHORS  
Strausberg, R.  
TITLE  
Submitted (03-NOV-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
JOURNAL

REMARK  
COMMENT  
NHI-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcaps-remail.nih.gov](mailto:gcaps-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalon@bcm.tmc.edu.  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the T.M.A.G.E. Consortium/BLIN at: <http://image.llnl.gov>  
Series: IRK Plate: 3 Row: e Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA g1: 3115203  
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers  
1..1716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2900081"  
/issue\_type="Placenta, choriocarcinoma"  
/clone\_id="NH\_MGC\_10"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
BASE COUNT 474 a 549 c 427 g 266 t  
ORIGIN

Query Match 71.1% Score 1656.8; DB 11; Length 1716;  
Best Local Similarity 99.9%; Pred. No. 1.9e-287;

Matches 1658; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

4 GGGATCTAGCATGTCGTGGAGAGCGGCGCTGGAGAGATTTGAAATCTGCCAGTA 63.  
57 GGGATCTAGCATGTCGTGGAGAGCGGCGCTGGAGAGATTTGAAATCTGCCAGTA 116  
64 ATGGGGATGGAGTAAAGCTTCCAAATGTCTCAACACAGGCGCCCAAGCCAG 123  
117 ATGGGGATGGAGTAAAGCTTCCAAATGTCTCAACACAGGCGCCCAAGCCAG 176  
124 CCAGCCACGACGACGACCAACCCCGCCCGCCAGAGACCTCAACCTTAACAAGCC 183  
177 CCAGCCACGACGACGACCAACCCCGCCCGCCAGAGACCTCAACCTTAACAAGCC 236  
184 AAGAGGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACATATGAA 243  
237 AAGAGGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACATATGAA 296  
244 CACGATTTGATGAGCTTCCAGAGCGCTGTGGATGCCGTCAAGCTACCTCCAT 303  
297 CACGATTTGATGAGCTTCCAGAGCGCTGTGGATGCCGTCAAGCTACCTCCAT 356  
304 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 363  
357 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 416  
364 AACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 423  
417 AACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 476  
424 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 483  
477 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 536  
484 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 543  
537 TACATTAAGATCTTAAAGCGCTATGATGATGGAACAATAAAGAGCGCTGGAAC 596  
544 AAGAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 603  
597 AAGAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 656  
604 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 663  
657 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 716  
664 CCGTGTGACAGGCGCTCAAGCGCTTCCGCGGTGACCGCGGAGCTCATCGTCAGAC 723  
|||||

717 CCGTGTGACAGGCGCTCAAGCGCTTCCGCGGTGACCGCGGAGCTCATCGTCAGAC 776  
724 CCGTGTGACAGGCGCTCAAGCGCTTCCGCGGTGACCGCGGAGCTCATCGTCAGAC 783  
777 CCGTGTGACAGGCGCTCAAGCGCTTCCGCGGTGACCGCGGAGCTCATCGTCAGAC 836  
784 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 843  
837 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 896  
844 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 903  
897 GCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 956  
904 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 963  
957 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 1016  
964 AAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1023  
1017 AAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1076  
1024 GACTCTGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1083  
1077 GACTCTGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1136  
1084 TGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1143  
1137 TGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1196  
1144 TACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1203  
1197 TACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1256  
1204 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 1263  
1257 CCAAGGACGACGACCAACCACTGCAATACGTCTCAGATGGTGTCTCAAGACAT 1316  
1264 GAGTTGTGCTGACGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1323  
1317 GAGTTGTGCTGACGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1376  
1324 CATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1383  
1377 CATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1436  
1384 ATGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1443  
1437 ATGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1496  
1444 ACCAAGGTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1503  
1497 ACCAAGGTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1556  
1504 GACAGTTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563  
1557 GACAGTTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1616  
1564 CAGCTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1623  
1617 CAGCTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1676  
1624 AAGAAAAAGGAG 1683  
1677 AAGAAAAAGGAG 1716  
|||||

RESULT 2  
BC030158 1743 bp mRNA linear HNC 20-MAY-2002  
LOCUS  
DEFINITION Homo sapiens, similar to bromodomain-containing 4, clone  
IMAGE:5086929, mRNA.  
ACCESSION BC030158



VERSION BC030158.1 GI:20988807  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1743)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland.  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Boufield, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAL Plate: 39 Row: n Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7657217  
This clone has the following problem: frame shifted.  
location/Qualifiers  
1..1743  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5086929"  
/tissue\_type="Liver, hepatocellular carcinoma"  
/clone\_id="NIH\_MGC\_100"  
/lab\_host="DH10B"  
/note="Vector: pOTB7"  
BASE COUNT 487 a 356 c 437 g 263 t  
ORIGIN  
Query Match 70.4%; Score 1640; DB 11; Length 1743;  
Best Local Similarity 99.4%; Pred. No. 1.9e-284;  
Matches 1646; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 4 GGGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTGAGAAATCTGCCACTA 63  
DB 87 GGGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTGAGAAATCTGCCACTA 146  
QY 64 ATGGGGATGAGCTAGAACTTCCCAATATGCTTACACACAGGCCCGGAGCCCAACCCAG 123  
DB 147 ATGGGGATGAGCTAGAACTTCCCAATATGCTTACACACAGGCCCGGAGCCCAACCCAG 206  
QY 124 CCAAGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183  
DB 207 CCAGCCAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 266  
QY 184 AAGAGGAGAGCAACCAATGCAATACCTGCTCAGAGTGTGCTCAAGACACTATGAAA 243  
DB 267 AAGAGGAGAGCAACCAATGCAATACCTGCTCAGAGTGTGCTCAAGACACTATGAAA 326  
QY 244 CACCAAGTTTGATGGCTTTCCAGCAGCCTGTGATGCCGTCAGCTGAACCTCCCTGAT 303  
|||||

DB 327 CACCAAGTTTGATGGCTTTCCAGCAGCCTGTGATGCCGTCAGCTGAACCTCCCTGAT 386  
QY 304 TACATTAAGATCATTAATAAGCCCTATGATATGGAACCAATTAAGACGGCTTGGAAAC 363  
DB 387 TACATTAAGATCATTAATAAGCCCTATGATATGGAACCAATTAAGACGGCTTGGAAAC 446  
QY 364 AACTATTACTGGATGCTCAGGAATGTATCCAGAGCTTCAACACTATGTTTACAAATTGT 423  
DB 447 AACTATTACTGGATGCTCAGGAATGTATCCAGAGCTTCAACACTATGTTTACAAATTGT 506  
QY 424 TACATCTACACAAAGCCCTGGAGATGATACATATGCTTAATAGGAGAAAGCTCTGAAAAGCTC 483  
DB 507 TACATCTACACAAAGCCCTGGAGATGATACATATGCTTAATAGGAGAAAGCTCTGAAAAGCTC 566  
QY 484 TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAAACCGAGATCATATATGATGTCAGGCA 543  
DB 567 TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAAACCGAGATCATATATGATGTCAGGCA 626  
QY 544 AAAGGAGAGAGAGCTGGAGAGAAAGAAACAGGGAGACGCAAAAACCTGGGCTTCCAGGTA 603  
DB 627 AAAGGAGAGAGAGCTGGAGAGAAAGAAACAGGGAGACGCAAAAACCTGGGCTTCCAGGTA 686  
QY 604 CCAAAACAAGCTCAAGCATTCGACTCTCCGAGACAGCCAGACCCCTCAGACCCCAATCTCTCT 663  
DB 687 CCAAAACAAGCTCAAGCATTCGACTCTCCGAGACAGCCAGACCCCTCAGACCCCAATCTCTCT 746  
QY 664 CCTGTGACAGGACAGCCTCACCCCTTCCCTGCGCCGTCACCCGAGACCTCATGCTCCAGAC 723  
DB 747 CCTGTGACAGGACAGCCTCACCCCTTCCCTGCGCCGTCACCCGAGACCTCATGCTCCAGAC 806  
QY 724 CCTGTGATGACAGTGTGCTCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 783  
DB 807 CCTGTGATGACAGTGTGCTCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 866  
QY 784 CCACAAACCCCAACCCGCTCAGCTCCCAAGCCGCTGACAGAGCAGCAGCAGCAGCAGCAGCAG 843  
DB 867 CCACAAACCCCAACCCGCTCAGCTCCCAAGCCGCTGACAGAGCAGCAGCAGCAGCAGCAGCAG 926  
QY 844 GCCACCCCAACCCGCTGTAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 903  
DB 927 GCCACCCCAACCCGCTGTAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 986  
QY 904 CCCACCAACCATTTGACAGCCATTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 963  
DB 987 CCCACCAACCATTTGACAGCCATTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1046  
QY 964 AAGCTGGGACAGCGCGGAG 1023  
DB 1047 AAGCTGGGACAGCGCGGAG 1106  
QY 1024 GACTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1083  
DB 1107 GACTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1166  
QY 1084 TGCAGCGGCATCTCAAGAGATGTTTGCAGAGACAGCGCGCTCAGCAGCAGCAGCAGCAGCAG 1143  
DB 1167 TGCAGCGGCATCTCAAGAGATGTTTGCAGAGAGACAGCGCGCTCAGCAGCAGCAGCAGCAGCAG 1226  
QY 1144 TACAAGCTGTGAGAGCTGAGAGCAGCTGAGGCTTACACAGCTACTGTGACATATCAAGAC 1203  
DB 1227 TACAAGCTGTGAGAGCTGAGAGCAGCTGAGGCTTACACAGCTACTGTGACATATCAAGAC 1286  
QY 1204 CCAATGAGCATGAGCAGACAAATCAAGCTCAATGAGAGCGCGCTGATACCGTATGATGCTCAG 1263  
DB 1287 CCAATGAGCATGAGCAGACAAATCAAGCTCAATGAGAGCGCGCTGATACCGTATGATGCTCAG 1346  
QY 1264 GAGTTTGGTGTGAGCTGAGCTCGATGATGTTTCAACTGCTATTAAGTCAACACCTCTGAC 1323  
DB 1347 GAGTTTGGTGTGAGCTGAGCTCGATGATGTTTCAACTGCTATTAAGTCAACACCTCTGAC 1406  
QY 1324 CATGAGTGTGTGAGCTGAGCTCGATGATGTTTCAACTGCTATTAAGTCAACACCTCTGAC 1383  
DB 1407 CATGAGTGTGTGAGCTGAGCTCGATGATGTTTCAACTGCTATTAAGTCAACACCTCTGAC 1466  
|||||

QY 1384 ATCCGGAGAGCCTGAGAGAGCCAGTGTGGCCGCTGTCTCCCGGAGTGGCCCTCC 1443  
 DB 1467 ATCCGGAGAGCCTGAGAGAGCCAGTGTGGCCGCTGTCTCCCGGAGTGGCCCTCC 1526  
 QY 1444 ACCAAGTGTGTGGCCCGCCCTCATCCAGAGAGAGAGAGAGAGAGTGTGGAGACT 1503  
 DB 1527 ACCAAGTGTGTGGCCCGCCCTCATCCAGAGAGAGAGAGAGAGTGTGGAGACT 1586  
 QY 1504 GACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563  
 DB 1587 GACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1646  
 QY 1564 CAGCTCAAGCCCTGAG 1623  
 DB 1647 CAGCTCAAGCCCTGAG 1706  
 QY 1624 AAGAAAAG 1659  
 DB 1707 AA 1742

RESULT 3  
 LOCUS BC011541 1776 bp mRNA linear HTC 07-AUG-2002  
 DEFINITION Mus musculus, similar to bromodomain-containing 4, clone  
 IMAGE:2648506, mRNA.  
 ACCESSION BC011541  
 VERSION BC011541.1 GI:15079397  
 KEYWORDS HTC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
 1 (bases 1 to 1776)  
 Strausberg, R.  
 Direct Submission  
 Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: amg@bcm.tmc.edu  
 Guarnatone, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,  
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
 Richards, S., Gibbs, R.A.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REMARK  
 COMMENT

FEATURES  
 source  
 1. 1776  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="IMAGE:2648506"  
 /tissue\_type="Mammary tumor. Metallothionin-TGF alpha  
 model. 10 month old virgin mouse. Taken by biopsy."  
 /clone\_lib="NCI CGAP\_Mam1"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
 Series: IRK Plate: 5 Row: j Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 10048443  
 This clone has the following problem: incomplete processing.  
 Location/Qualifiers

BASE COUNT 545 a 517 c 401 g 313 t  
 ORIGIN  
 Query Match 60.6%; Score 1411; DB 11; Length 1776;  
 Best Local Similarity 90.3%; Pred. No. 2,4e-243;  
 Matches 1520; Conservative 0; Mismatches 160; Indels 3; Gaps 1;  
 4 GGGATCACTAGATGTCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
 DB 90 GGGATCACTAGATGTCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
 QY 64 ATGGGGATGAGACTAGAACTTCCCAATGTCTACAAACACAGAGGCCCAAGCCAG 123  
 DB 150 ATGGGGATGAGACTAGAACTTCCCAATGTCTACAAACACAGAGGCCCAAGCCAG 209  
 QY 124 CCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183  
 DB 210 CCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 269  
 QY 184 AAGAGCAGACCAACCACTGCAATACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243  
 DB 270 AAGAGCAGACCAACCACTGCAATACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 329  
 QY 244 CACCACTTTGATGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 303  
 DB 330 CACCACTTTGATGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389  
 QY 304 TACTATAGATGATTAAG 363  
 DB 390 TACTATAGATGATTAAG 449  
 QY 364 AACTATTAAGATGATTAAG 423  
 DB 450 AACTATTAAGATGATTAAG 509  
 QY 424 TACTATTAAGATGATTAAG 483  
 DB 510 TACTATTAAGATGATTAAG 569  
 QY 484 TTTCTGCAAAATTAATAGTACCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 543  
 DB 570 TTTCTGCAAAATTAATAGTACCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629  
 QY 544 AAG 603  
 DB 630 AAG 689  
 QY 604 CCAAAACAACTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
 DB 690 CCAAAACAACTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 749  
 QY 661 CTTCTGTGAG 720  
 DB 750 CTTCTGTGAG 809  
 QY 721 ACCCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 DB 810 ACCCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 869  
 QY 781 CAGCCCAACCCCAAG 840  
 DB 870 CAGCCCAACCCCAAG 929  
 QY 841 GGGGCAACCCCAAG 900  
 DB 930 GGGGCAACCCCAAG 989  
 QY 901 ACCCCCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
 DB 990 ACCCCCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1049  
 QY 961 ACCAAGTGGGAG 1020





REFERENCE 1 (bases 1 to 893)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: LLM13225 Row: d Column: 24  
 High quality sequence stop: 675.  
 Location/Qualifiers  
 1..893  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="6020711"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epithelioid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."

BASE COUNT 259 a 298 c 199 g 137 t

ORIGIN

Query Match 33.4%; Score 779; DB 14; Length 893;  
 Best Local Similarity 95.3%; Pred. No. 6e-130;  
 Matches 826; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 200 AACGCAATACCTGCTCAGAGTGGTGCACAGACATATGGAACACAGTTGCATGCG 259  
 Db 1 AACTGCATACCTGCTCAGAGTGGTGCACAGACATATGGAACACAGTTGCATGCG 60  
 QY 260 CTTTCAGCAGCCCTGTGATGCGCTCAAGCTGAACTCCCTGATTACTATTAAGATCATTA 319  
 Db 61 CTTTCAGCAGCCCTGTGATGCGCTCAAGCTGAACTCCCTGATTACTATTAAGATCATTA 120  
 QY 320 AAAGCCTATGATATGGGAACATTAAGAAGCCCTGGAAAAACACTATTACTGGAATG 379  
 Db 121 AAAGCCTATGATATGGGAACATTAAGAAGCCCTGGAAAAACACTATTACTGGAATG 180  
 QY 380 CTCAGATGATATGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 439  
 Db 181 CTCAGATGATATGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 240  
 QY 440 CTGAGATGATATGATCTTATATGCGAGAAAGCTCTGAAAAGCTCTTCTGCAAAAAATTA 499  
 Db 241 CTGAGATGATATGATCTTATATGCGAGAAAGCTCTGAAAAGCTCTTCTGCAAAAAATTA 300  
 QY 500 ATGAGCTATCCACAGAAAGAAACGAGATCATGATATGTCAGGCAAAAGAAAGAGAGCGTG 559  
 Db 301 ATGAGCTATCCACAGAAAGAAACGAGATCATGATATGTCAGGCAAAAGAAAGAGAGCGTG 360  
 QY 560 GGAAGAAAGAAAGAGGAGACAAACCTGGCGTTTCCAGGTTACCAAAACCAACTCAAG 619  
 Db 361 GGAAGAAAGAAAGAGGAGACAAACCTGGCGTTTCCAGGTTACCAAAACCAACTCAAG 420  
 QY 620 CATGACTCTCTCGCAGACCCAGACCCCTCAGCGAAATCTCTCTGTGTGAGGCGCAAGC 679  
 Db 421 CATGACTCTCTCGCAGACCCAGACCCCTCAGCGAAATCTCTCTGTGTGAGGCGCAAGC 480  
 QY 680 CTCACCCCTTCTGCGCGTCAACCCGAGACTCATGTCACAGACCCCTGTATGACAGTGG 739  
 Db 481 CTCACCCCTTCTGCGCGTCAACCCGAGACTCATGTCACAGACCCCTGTATGACAGTGG 540  
 QY 740 TGCTTCCCAAGCCACTGAGAGCGCCCGCAGTGGCCCGCCAGCCCAACCCCGACCGG 799  
 Db 541 TGCTTCCCAAGCCACTGAGAGCGCCCGCAGTGGCCCGCCAGCCCAACCCCGACCGG 600

QY 800 CTCGAGCTCCCCAGCCCTTACAGAGCCACCCACCATCATGCGGCGCACCCAGAGCCTG 859  
 Db 601 CTCGAGCTCCCCAGCCCTTACAGAGCCACCCACCATCATGCGGCGCACCCAGAGCCTG 660  
 QY 860 TGAAGACAAAGAGGAGTGAAGAGAAAGAGAGACACCCACCCACCATTTGAC 919  
 Db 661 TGAAGACAAAGAGGAGTGAAGAGAAAGAGAGACACCCACCCACCATTTGAC 720  
 QY 920 CCATTTCAGACCCACCCCTGCTGCCCCCGGAGCCCAAGCAAGCTGGGCCAGCGG 979  
 Db 721 CCATTTCAGACCCACCCCTGCTGCCCCCGGAGCCCAAGCAAGCTGGGCCAGCGG 780  
 QY 980 GGGAGAG---CAGCGGCTGTGAAA---CCTCCAAAGAGAGAGCTGCTCAGC 1033  
 Db 781 CGGGAGAGAGAGCGCGGCGCTGTGAAAACCTCCCAAGAAAGAGAGCTGCTCAGC 840  
 QY 1034 AGCACCCAGCACAGAGAGAGAGCA 1060  
 Db 841 AGCACCCAGCACAGAGAGAGAGAA 867

RESULT 7  
 BQ214403 933 bp mRNA linear EST 02-MAY-2002  
 LOCUS AGENCOURT\_7589671 NIH\_MGC\_92 Homo sapiens CDNA clone IMAGE:6067786  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ214403  
 VERSION BQ214403.1 GI:20395803  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 933)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: LLM13347 Row: n Column: 11  
 High quality sequence stop: 659.  
 Location/Qualifiers  
 1..933  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="6067786"  
 /clone\_lib="NIH\_MGC\_92"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: testis; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dt primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 288 a 304 c 196 g 145 t

ORIGIN

Query Match 33.1%; Score 770.2; DB 14; Length 933;  
 Best Local Similarity 92.3%; Pred. No. 2.3e-128;  
 Matches 846; Conservative 0; Mismatches 63; Indels 8; Gaps 3;

QY 25 GAGAGCGCCCTGGGAGAGATGGAATCTGCAGTAATGGGGGATGGAGTGAAGT 84  
 Db 13 GCGAGCGCCCTGGGAGAGATGGAATCTGCAGTAATGGGGGATGGAGTGAAGT 72

```
OY 85 TCCCAATGCTCTACACACAGGCGCCAGCCCAACCCAGCCAGCCAAAGCCAGCAGCAGC 144
    |||||||
Db 73 TCCCAATGCTCTACACACAGGCGCCAGCCCAACCCAGCCAGCCAAAGCCAGCAGCAGC 132
OY 145 AACCCCCCGCCCGAGAGAGCTCCAAACCTTAACAAGCCCAAGAGGAGCAGCAACCACTG 204
    |||||||
Db 133 AACCCCCCGCCCGAGAGAGCTCCAAACCTTAACAAGCCCAAGAGGAGCAGCAACCACTG 132
OY 205 CAATACCTGCTCAGAGTGTGCTCAAGACACTATGGAAGAACACAGTTTGATGCGCTTTC 264
    |||||||
Db 193 CAATACCTGCTCAGAGTGTGCTCAAGACACTATGGAAGAACACAGTTTGATGCGCTTTC 252
OY 265 CACAGCCTGTGTGATGCGCTCAAGCTGAACCTCCGATTAATTAAGATCAATTAAGC 324
    |||||||
Db 253 CACAGCCTGTGTGATGCGCTCAAGCTGAACCTCCGATTAATTAAGATCAATTAAGC 312
OY 325 CCTATGATATGGAACCAATTAAGAACCGCTTGAAACCAACTATTACTGGAATGCTCAG 384
    |||||||
Db 313 CCTATGATATGGAACCAATTAAGAACCGCTTGAAACCAACTATTACTGGAATGCTCAG 372
OY 385 GAATGATCCAGAGCTTCAACACTATGTTTACAATTTGTACATCTAACCAAGCCTGGA 444
    |||||||
Db 373 GAATGATCCAGAGCTTCAACACTATGTTTACAATTTGTACATCTAACCAAGCCTGGA 432
OY 445 GATGACATAGTCTTAATGGGAGAGCTGGAAGAGCTCTTCTGCAAAAAATTAATAG 504
    |||||||
Db 433 GATGACATAGTCTTAATGGGAGAGCTGGAAGAGCTCTTCTGCAAAAAATTAATAG 492
OY 505 CTACCCACAGAAAGAACCCAGATCATATGATGTCAGGCAAAAGAAAGAGAGCTGGAG 564
    |||||||
Db 493 CTACCCACAGAAAGAACCCAGATCATATGATGTCAGGCAAAAGAAAGAGAGCTGGAG 552
OY 565 AAAGAAACAGGAGACAGAAAACCTGGCGTTTCCACGGTACCAAAACAACTCAAGCATCG 624
    |||||||
Db 553 AAAGAAACAGGAGACAGAAAACCTGGCGTTTCCACGGTACCAAAACAACTCAAGCATCG 612
OY 625 ACACCTCCGAGAGCCAGACCCCTCAGCGGAATCTCTCTGAGAGGCGAGCCCTCAG 684
    |||||||
Db 613 ACACCTCCGAGAGCCAGACCCCTCAGCGGAATCTCTCTGAGAGGCGAGCCCTCAG 672
OY 685 CCGTTCCCTGCGCTCAACCCCGGACCTCATGCTCAGACCCCTGATGAGAGTGTGCT 744
    |||||||
Db 673 CCGTTCCCTGCGCTCAACCCCGGACCTCATGCTCAGACCCCTGATGAGAGTGTGCT 732
OY 745 CCGCAGCAGCTGAGAGCGCCCGCCAG--TGCCTCCCAAGCCCAACCCCAAGCCTC 802
    |||||||
Db 733 CCGCAGCAGCTGAGAGCGCCCGCCAGGTCGCGCCCAAGCAACCCCAAGCCTC 792
OY 803 C-AGCTCCCAAGCGGATACAGAGCCAGCCACCC-----ATCATCGGGGCGCAGCCAGC 856
    |||||||
Db 793 CAAGCTCCCAAGCGGATACAGAGCCAGCCAGCCATCATTTGCGGGGCGCAGCCCAAGC 852
OY 857 CTGTGAAGACAAAGAGGAGTGAAGAGAAAGCAGACACACCCCGCAGCAGCATTTG 916
    |||||||
Db 853 CTGTGGGAGAGACCAAAAAAGGAGTGAAGAAAGGAAAAAGCCCAAAAAATCAACCA 912
OY 917 ACCCCATTCACGAGCCA 933
    |||||||
Db 913 TCCCGCCAGCCAGCCCA 929
```

```
RESULT 8
LOCUS B0653187 975 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCODRT_8341467 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267841
5', mRNA sequence.
ACCESSION B0653187
VERSION B0653187.1 GI:21777359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

```
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
        Email: cgrabs-remail.nih.gov
        Tissue Procurement: CGAP (Stanford)
        CDNA Library Preparation: Rubin Laboratory
        CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.lnl.gov
        Plate: LLM2438 row: n column: 02
        High quality sequence stop: 656.
        Location/Qualifiers
            1..975
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6267841"
                /clone_1lb="NIH_MGC_100"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; CDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GCCACGAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-CDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                library."
BASE COUNT 275 a 304 c 246 g 145 t 5 others
ORIGIN
Query Match 32.5%; Score 756.6; DB 14; Length 975;
Best Local Similarity 93.7%; Pred. No. 6.2e-126;
Matches 841; Conservative 0; Mismatches 49; Indels 8; Gaps 5;
OY 4 GGGATCACTAGCATGTCTGGGAGAGCGGCCCTGGGAGAGATTTGAAATCTGCCAGTA 63
    |||||||
Db 77 GGGATCACTAGCATGTCTGGGAGAGCGGCCCTGGGAGAGATTTGAAATCTGCCAGTA 136
OY 64 ATGGGGAGATGACTAGAAATCTCCAAATGTCTTACAACACAGGCCCAAGCCCAAG 123
    |||||||
Db 137 ATGGGGAGATGACTAGAAATCTCCAAATGTCTTACAACACAGGCCCAAGCCCAAG 196
OY 124 CCAAGCCAGGACGACGACCAACCCCGCGCCCGCCAGAGACCTCCAAACCTTAAGAGCC 183
    |||||||
Db 197 CCAAGCCAGGACGACGACCAACCCCGCGCCCGCCAGAGACCTCCAAACCTTAAGAGCC 256
OY 184 AAGAGCAGACCAACCAACTGCAATACCTGCTCAGAGTGTGCTCAAGACATATGAAA 243
    |||||||
Db 257 AAGAGCAGACCAACCAACTGCAATACCTGCTCAGAGTGTGCTCAAGACATATGAAA 316
OY 244 CACCACTTTGATAGGCTTTCCAGCAGCTGTGGAATGCCGTCAAGCTGAACCTCCTGAT 303
    |||||||
Db 317 CACCACTTTGATAGGCTTTCCAGCAGCTGTGGAATGCCGTCAAGCTGAACCTCCTGAT 376
OY 304 TACTATAAGATCAATTAAGAGGCTATGATATGGAAGAAATAAGAAAGGCTTGAAAC 363
    |||||||
Db 377 TACTATAAGATCAATTAAGAGGCTATGATATGGAAGAAATAAGAAAGGCTTGAAAC 436
OY 364 AACTATTACTGGAATGCTCAGAAATATCAAGAGCTTCAACACTATGTTTACAATTTGT 423
    |||||||
Db 437 AACTATTACTGGAATGCTCAGAAATATCAAGAGCTTCAACACTATGTTTACAATTTGT 496
OY 424 TACATCTACAACAGCCTGAGATGACATAGTCTTAATGAGCAGAACTCTGGAAGAGCTC 483
    |||||||
Db 497 TACATCTACAACAGCCTGAGATGACATAGTCTTAATGAGCAGAACTCTGGAAGAGCTC 556
    |||||||
OY 484 TTCTTCAAAAAATTAATGAGCTACCCACAGAAAGAAAGGAGATCTGATAGTCCAGGCA 543
    |||||||
```

DB	557	TCCTGCAAAAATAAATAGAGCTACCACAGAAAGAAACCGACATCATGTATGTCAGGCA	616
QY	544	AAAGGAAGAGGACGTCGGGAGCAAAAGAAACAGGAGACGACAAACCTGGGGCTTTCCAGGTA	603
Db	617	AAAGGAAGGAGCGTCGGGAGGAAAGAAACAGGAGACAGCANAACCTGGGCTTTCCAGGTA	676
QY	604	CCAAACACAACTCAAGCATCGACTCTCCGGAGACCCAGACCCCTTCAGCGAATCTCTCT	663
Db	677	CCAAACACAACTCAAGCATCGACTCTCCGGAGACCCAGACCCCTTCAGCGAATCTCTCT	736
QY	664	CTGTGCAAGGCGACGCTCACCCCTTCCCTGGCGCTCACCCCGGAGCTCATCTGTCAGACG	723
Db	737	CCTGTGCAAGGCGACGCTCACCCCTTCCCTGGCGCTCACCCCGGAGCTCATCTGTCAGACG	796
QY	724	CCTGTGAT- GACAGTGTGTCTCTCCAGCAGCT---GAGAGAGCCCGCGCAGTCCGCC	779
Db	797	CCTGTGATGAGACAGTGTGTCTCTCCAGCAGCT---GAGAGAGCCCGCGCAGTCCGCC	856
QY	780	CCAGGCA-CAACCCCGACCCGCTCCAGCTCCCG--AGCCCGTACAGAGCAGCAGCA-CCCA	835
Db	857	CCAGGCAACACCCCGACCCCGCTCCAGCTCCCG--AGCCCGTACAGAGCAGCAGCAGCA-CCCA	916
QY	836	TCATGCGGCGCCACCCACAGCCTGTGAGAGACAAAGAGGAGTGAAGAGAGAGAGA	893
Db	917	TCATGCGGCGCGCCCGCCCGCAGCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGA	974
RESULT 9			
BO645818		925 bp	mrna linear EST 15-JUL-2002
LOCUS			
DEFINITION	AGENCOURT_8350724 NIH_MGC_100 Homo sapiens	cdna clone	IMAGE:62871898
ACCESSION	BO645818		
VERSION	BO645818.1	GI:21769990	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 925)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgsabbs-remail.nih.gov		
	Tissue Procurement: CGAP (Starford)		
	CDNA Library Preparation: Rubin Laboratory		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/MLN at:		
	http://image.llnl.gov		
	Plate: LICM2489 row: d column: 06		
FEATURES	High quality sequence stop: 684.		
Source	Location/Qualifiers		
	1..925		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6287189"		
	/clone_lib="NIH_MGC_100"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: liver; Vector: pOTR1; Site_1: XhoI; Site_2:		
	EcoRI; CDNA made by oligo-dT priming. Directionally cloned		
	into EcoRI/XhoI sites using the following 5' adaptor:		
	GGCAGCAG(G). Size-selected >500bp for average insert size		
	1.8kb. Library constructed by Ling Hong in the laboratory		
	of Gerald M. Rubin (University of California, Berkeley)		
	using ZAP-CDNA synthesis kit (Stratagene) and Superscript		
	II RT (Life Technologies). Note: this is a NIH_MGC		
	Library."		
BASE COUNT	261 a 293 c 220 g 147 t 2 others		
ORIGIN			

Query Match	Similarity	32.3%	Score 752.47	DB 14	Length 925
Best Local	Similarity	97.0%	Pred. No. 3,5e-125		
Matches	776	Conservative	0	Mismatches	23
				Indels	1
				Gaps	1
QY	4	GGGATCACTAGCATGTCTGGCGGAGAGCGGCCCTGGGACGAGATTGAAATCTGCCAGTA	63		
Db	77	GGGATCACTAGCATGTCTGGCGGAGAGCGGCCCTGGGACGAGATTGAAATCTGCCAGTA	136		
QY	64	ATGGGGGATGGGACTAGAAAATCTTCCCAATGTGTACAAACACAGGCCAGGCCCAACCCAG	123		
Db	137	ATGGGGGATGGGACTAGAAAATCTTCCCAATGTGTACAAACACAGGCCAGGCCCAACCCAG	196		
QY	124	CCAGCCAAACGCGAGCGAGCACAACCCCGCCGCCCGAGAGACTCCAAACCTTAACAGGCC	183		
Db	197	CCAGCCAAACGCGAGCGAGCACAACCCCGCCGCCCGAGAGACTCCAAACCTTAACAGGCC	256		
QY	184	AAGAGGCGAGCCAAACCACTGCAATACCTGCTCAGAGTGGTCTCAAGACATATGGAAA	243		
Db	257	AAGAGGCGAGCCAAACCACTGCAATACCTGCTCAGAGTGGTCTCAAGACATATGGAAA	316		
QY	244	CACCGATTGGCATGGCCCTTTCCAGACGCTGTGATGCCGTAACTGAACCTCCCTGAT	303		
Db	317	CACCGATTGGCATGGCCCTTTCCAGACGCTGTGATGCCGTAACTGAACCTCCCTGAT	376		
QY	304	TACTATTAGATATATTAATAAAGCCCTTATGGATATGGGAAACAATTAAGAGGGCTTGGAAA	363		
Db	377	TACTATTAGATATATTAATAAAGCCCTTATGGATATGGGAAACAATTAAGAGGGCTTGGAAA	436		
QY	364	AACATATTACTGGAAATGCTCAGGAATGTATCCAGAGCTTCMACATATGTTTACAATGT	423		
Db	437	AACATATTACTGGAAATGCTCAGGAATGTATCCAGAGCTTCMACATATGTTTACAATGT	496		
QY	424	TACATCTCAACAACCGCTGGAGATACATAGCTTTAATGGCAAGGCTTGGAAAAAGCTC	483		
Db	497	TACATCTCAACAACCGCTGGAGATACATAGCTTTAATGGCAAGGCTTGGAAAAAGCTC	556		
QY	484	TTCTTGCAAAAAATTAATGAGGTACCCACAGAAAGAACCGAGATCATGTATATCCAGGCA	543		
Db	557	TTCTTGCAAAAAATTAATGAGGTACCCACAGAAAGAACCGAGATCATGTATATCCAGGCA	616		
QY	544	AAAGGAGAGGAGCGTGGGAGGAAAGAAACAGGAGACGCAAAAACCTGGCGTTTCCAGGTA	603		
Db	617	AAAGGAGAGGAGCGTGGGAGGAAAGAAACAGGAGAGGCAAAAACCTGGCGTTTCCAGGTA	676		
QY	604	CCAAACACAACCTCAAGCATCGACTCTCCGCGAGACCCAGACCCCTTCAGGATCTCTCT	663		
Db	677	CCAAACACAACCTCAAGCATCGACTCTCCGCGAGACCCAGACCCCTTCAGGATCTCTCT	736		
QY	664	CCTGTGCAAGGCGACGCTCAACCTCTCCCTGCGCTCACCCCGGAGCTCATGTCCAGACC	723		
Db	737	CCTGTGCAAGGCGACGCTCAACCTCTCCCTGCGCTCACCCCGGAGCTCATGTCCAGACC	796		
QY	724	CCTGTGATGAC-AGTGTGCTCTCCCGAGCGACTGAGAGGCCCGCGCAGATGCCGCCCA	782		
Db	797	CCTGTGATGACAGTGTGCTCTCCCGAGCGACTGAGAGGCCCGCGCAGATGCCGCCCA	856		
QY	783	GGCACAACCCCAACCCGCTC 802			
Db	857	CCAAACCAACCCCAACCCGCTC 876			
RESULT 10					
LOCUS	B0647311	899 bp	mrna	linear	EST 15-JUL-2002
DEFINITION	AGNCNCURT_8350802 NIH_MGC_100 Homo sapiens cdna IMAGE:6287170				
ACCESSION	B0647311				
VERSION	B0647311.1	GI:21771483			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				



REFERENCE 1 (bases 1 to 899)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM2489 row: c column: 11  
High quality sequence stop: 680.  
Location/Qualifiers

FEATURES  
source

1. 899  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6287170"  
/clone\_lib="NIH-MGC.100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 257 a 278 c 217 g 144 t 3 others  
ORIGIN

Query Match 31.6%; Score 736.2; DB 14; Length 899;  
Best Local Similarity 95.4%; Pred. No. 2.8e-122;

Matches 768; Conservative 0; Mismatches 35; Indels 2; Gaps 1;

4 GGGATCCTAGCATGTCTCGGAGAGCGGCGCTGGGACGAGATTGGAATCTGCCAGTA 63  
77 GGGATCCTAGCATGTCTCGGAGAGCGGCGCTGGGACGAGATTGGAATCTGCCAGTA 136  
64 ATGGGGGATGACTAGAAAATCTCCAAATGTCTACAAACAGGCCGCCAACCAGCC 123  
137 ATGGGGGATGACTAGAAAATCTCCAAATGTCTACAAACAGGCCGCCAACCAGCC 196  
124 CGAGCCAAAGCAGCAGCAACACCCCGCGCCCGCAGAGACCTCCAAACCTTAACAGCC 183  
197 CGAGCCAAAGCAGCAGCAACACCCCGCGCCCGCAGAGACCTCCAAACCTTAACAGCC 256  
184 MAGAGCAGACCAACCACTGCAATACCTGCTCAGAGTGGTCTCAAGACATATGAAA 243  
257 MAGAGCAGACCAACCACTGCAATACCTGCTCAGAGTGGTCTCAAGACATATGAAA 316  
244 CACCAAGTTGCAATGGCTTCCAGACAGCTGTGGATCCGTCAGCTGAACCTCCCTGAT 303  
317 CACCAAGTTGCAATGGCTTCCAGACAGCTGTGGATCCGTCAGCTGAACCTCCCTGAT 376  
304 TACTATTAAGATCAATAAAGCGCTATGATGGGAAACAATAAAGAGCGCTTGAAAAC 363  
377 TACTATTAAGATCAATAAAGCGCTATGATGGGAAACAATAAAGAGCGCTTGAAAAC 436  
364 AACTATTAAGATCAATAAAGCGCTATGATGGGAAACAATAAAGAGCGCTTGAAAAC 423  
437 AACTATTAAGATCAATAAAGCGCTATGATGGGAAACAATAAAGAGCGCTTGAAAAC 496  
424 TACATCTACACAGCGCTGAGATGACATATGCTTAATGAGAGAGCGCTTGAAAACGTC 483  
497 TACATCTACACAGCGCTGAGATGACATATGCTTAATGAGAGAGCGCTTGAAAACGTC 556  
484 TTCTTGCAAAAAATAATGAGCTACCCCAAGAAAGAGAGAGATGATGATGCTCCAGCA 543

|||||  
Db 557 TTCTTGCAAAAAATAATGAGCTACCCACAGAGAAACCGAGATCAATGATCCAGGCA 616  
Oy 544 AAGAGAGAGAGAGCTGGAGAGAAAGAGAGAGCAAAACCTGGCTTCCAGGTA 603  
Db 617 AAGAGAGAGAGAGCTGGAGAGAAAGAGAGAGCAAAACCTGGCTTCCAGGTA 676  
Oy 604 CCAACACACATCAAGCATGATCTCTCCGACAGACCCAGACCCCTGAGCAATCTCT 663  
Db 677 CCAACACACATCAAGCATGATCTCTCCGACAGACCCAGACCCCTGAGCAATCTCT 736  
Oy 664 CCTGTGACAGCCAGCGCTTACCCCTTCCCTGCGGTACACCCGAGCTCATCTCCAGAC 723  
Db 737 CCTGTGACAGCCAGCGCTTACCCCTTCCCTGCGGTACACCCGAGCTCATCTCCAGAC 796  
Oy 724 CCTGTGATG-ACAGTGGTGTCTCCGACGACCTGACAGCGCCCGCAGTGCCTCC 781  
Db 797 CCTGTGATGACAGTGGTGTCTCCGACGACCCAGACCCCTGAGCAATCTCT 856  
Oy 782 AGCCACACACCCCGACCCGCTCCAGC 806  
Db 857 CCCCCCAGGCAAAACCCCGAC 881

RESULT 11  
B0650713 931 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT.8419166 NIH-MGC.100 Homo sapiens CDNA clone IMAGE:6284393  
DEFINITION 5', mRNA sequence.  
ACCESSION B0650713  
VERSION B0650713.1 GI:21774885  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 931)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM2481 row: c column: 18  
High quality sequence stop: 639.  
Location/Qualifiers

FEATURES  
source

1. 931  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6284393"  
/clone\_lib="NIH-MGC.100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 258 a 299 c 219 g 149 t 6 others  
ORIGIN

Query Match 31.6%; Score 735.8; DB 14; Length 931;  
Best Local Similarity 95.8%; Pred. No. 3.4e-122;



Matches 775; Conservative 0; Mismatches 31; Indels 3; Gaps 2;			
QY	4	GGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTTGAGAAATCTGCCAGTA	63
Db	77	GGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTTGAGAAATCTGCCAGTA	136
QY	64	ATGGGGATGAGCTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	123
Db	137	ATGGGGATGAGCTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	196
QY	124	CCAGCCAAAGGAGGAGCAGCAGCCACCCCGCCCGCCAGAGACTTCCAAACCTTCAAGGCC	183
Db	197	CCAGCCAAAGGAGGAGCAGCAGCCACCCCGCCCGCCAGAGACTTCCAAACCTTCAAGGCC	256
QY	184	AAGAGGAGAGCAACCAACTGATACCTGCTCAGAGTGGTCTCAAGACTATGAGAA	243
Db	257	AAGAGGAGAGCAACCAACTGATACCTGCTCAGAGTGGTCTCAAGACTATGAGAA	316
QY	244	CACCACTTTGATGGCTTTTCACAGAGCTGTGAGATGGCTCAAGCTGAACCTCCCTGAT	303
Db	317	CACCACTTTGATGGCTTTTCACAGAGCTGTGAGATGGCTCAAGCTGAACCTCCCTGAT	376
QY	304	TACTATAAGATCATTAAGAGCCTATGATATGGAGCAATAAAGAGCCTTGGAAAC	363
Db	377	TACTATAAGATCATTAAGAGCCTATGATATGGAGCAATAAAGAGCCTTGGAAAC	436
QY	364	AACATTTACTGAGATGCTCAGAGATGTATCCAGAGCTTCAACACTATGTTACAAATGT	423
Db	437	AACATTTACTGAGATGCTCAGAGATGTATCCAGAGCTTCAACACTATGTTACAAATGT	496
QY	424	TACATCTACAAACAAGCCTGAGATGACATAGTCTTAATGCGAGAGCTTGGAAAGCTC	483
Db	497	TACATCTACAAACAAGCCTGAGATGACATAGTCTTAATGCGAGAGCTTGGAAAGCTC	556
QY	484	TTCTTGGAAAAAATTAATGAGCTACGACAGAGAAACCGAGATCATGATGTCCAGCA	543
Db	557	TTCTTGGAAAAAATTAATGAGCTACGACAGAGAAACCGAGATCATGATGTCCAGCA	616
QY	544	AAAGGAAGAGAGCTGGGAGGAAAGAAACAGAGGACAGCAAAACCTGGCTTCCAGGTA	603
Db	617	AAAGGAAGAGAGCTGGGAGGAAAGAAACAGAGGACAGCAAAACCTGGCTTCCAGGTA	676
QY	604	CCAAACCAACTCAAGCATGCTCTCCGAGAGCCAGACCCCTCAAGCCCAATCTCTCT	663
Db	677	CCAAACCAACTCAAGCATGCTCTCCGAGAGCCAGACCCCTCAAGCCCAATCTCTCT	736
QY	664	CCGTGTGAGGCGACGCTCAACCCCTTCCCTGCGCTCAACCCGAGCTCATGTCTCAAGCC	723
Db	737	CCGTGTGAGGCGACGCTCAACCCCTTCCCTGCGCTCAACCCGAGCTCATGTCTCAAGCC	796
QY	724	CTGTGTATG-ACAGTGTGCTCCCGCAG--CAACTGAGAGAGCCCGCCAGTGCCCGCC	780
Db	797	CTGTGTATG-ACAGTGTGCTCCCGCAG--CAACTGAGAGAGCCCGCCAGTGCCCGCC	856
QY	781	CAGCCACACCCCGGCTCCAGCTCC 809	
Db	857	CCCCAGGAGCAACCCCGGCTCC 885	
RESULT 12			
LOCUS B0652326 919 bp mRNA linear EST 15-JUL-2002			
DEFINITION AGENCOUTRT_8350731 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6287333			
ACCESSION B0652326			
VERSION B0652326.1 GI:21776498			
KEYWORDS RST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1 (bases 1 to 919)			
AUTHORS NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .			

TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL		Unpublished (1999)	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: gcgaps-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LNCM2489 row: j column: 06 High quality sequence stop: 593. Location/Qualifiers	
FEATURES		1. 919 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6287333" /clone_1kb="NIH_MGC_100" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: Liver; Vector: pOTB7; Site-1: XhoI; Site-2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."	
BASE COUNT		259 a 289 c 221 g 149 t 1 others	
ORIGIN			
Query Match		31.5%; Score 733.4; DB 14; Length 919;	
Best Local Similarity		94.5%; Pred. No. 9.1e-122;	
Matches 793; Conservative 0; Mismatches 41; Indels 5; Gaps 3;			
QY	4	GGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTTGAGAAATCTGCCAGTA	63
Db	77	GGATCACTAGCATGTCTGCGAGAGCGCCCTGGAGAGATTTGAGAAATCTGCCAGTA	136
QY	64	ATGGGGATGAGCTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	123
Db	137	ATGGGGATGAGCTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	196
QY	124	CCAGCCAAAGGAGGAGCAGCAGCCACCCCGCCCGCCAGAGACTTCCAAACCTTCAAGGCC	183
Db	197	CCAGCCAAAGGAGGAGCAGCAGCCACCCCGCCCGCCAGAGACTTCCAAACCTTCAAGGCC	256
QY	184	AAGAGGAGAGCAACCAACTGATACCTGCTCAGAGTGGTCTCAAGACTATGAGAA	243
Db	257	AAGAGGAGAGCAACCAACTGATACCTGCTCAGAGTGGTCTCAAGACTATGAGAA	316
QY	244	CACCACTTTGATGGCTTTTCACAGAGCTGTGAGATGGCTCAAGCTGAACCTCCCTGAT	303
Db	317	CACCACTTTGATGGCTTTTCACAGAGCTGTGAGATGGCTCAAGCTGAACCTCCCTGAT	376
QY	304	TACTATAAGATCATTAAGAGCCTATGATATGGAGCAATAAAGAGCCTTGGAAAC	363
Db	377	TACTATAAGATCATTAAGAGCCTATGATATGGAGCAATAAAGAGCCTTGGAAAC	436
QY	364	AACATTTACTGAGATGCTCAGAGATGTATCCAGAGCTTCAACACTATGTTACAAATGT	423
Db	437	AACATTTACTGAGATGCTCAGAGATGTATCCAGAGCTTCAACACTATGTTACAAATGT	496
QY	424	TACATCTACAAACAAGCCTGAGATGACATAGTCTTAATGCGAGAGCTTGGAAAGCTC	483
Db	497	TACATCTACAAACAAGCCTGAGATGACATAGTCTTAATGCGAGAGCTTGGAAAGCTC	556
QY	484	TTCTTGGAAAAAATTAATGAGCTACGACAGAGAAACCGAGATCATGATGTCCAGCA	543
Db	557	TTCTTGGAAAAAATTAATGAGCTACGACAGAGAAACCGAGATCATGATGTCCAGCA	616

QY 544 AAAGGAAAGCAGCCTGGGAGGAAAAGAAACAGGGACAGCAAACTCTGGCTTTCCACGGTA 603

Db 617 AAAGGAAAGGAGCCTGGGAGGAAAATATACAGTACAGCAAACTCTGGCTTTCCACGGTA 676

QY 604 CCAAACACACTCAAGATCGACTCTCTCCGACAGACCCAGACCCCTTAGCCGAATCTCTCT 663

Db 677 CCAAACACAACTCAAGCATCGACTCTCTCCGACAGACCCAGACCCCTTAGCCGAATCTCTCT 736

QY 664 CCTGTGCAGGCGCAGCCTCACCCCTTCCCTGCGCGTACACCCCGGACTCATGTCTCCAGACC 723、

Db 737 CCTGTGCAGGCGCAGCCTCACCCCTTCTTCTGCGGTACACCCCGGACTCATGTCTCCAGACC 796

QY 724 CCTGTCAATG-ACAGTGTGTGCTCTCCCGAGCCAC--TTGCAGACGCGCCCGCGCAATGCCCGCC 780

Db 797 CCTGTCAATGACAGTGTGTGCTCTCCCGAGCCACCTTGAGACAGACCCCGCGCAATGGCGCCCG 856

QY 781 CAGGCA--CAACCGCCGACCGCTCGACGCTCCCGACCGCCGCTACAGAGCCACCCACCCATC 837

Db 857 CCACCAAGAAACCCCGACCCCGCTTACAGATCTCCCGACGCGCCGAGCTACAGACCCCGCCCGACC 915

RESULT	LOCUS	DEFINITION
13 B0650970	B0650970	926 bp mRNA linear EST 15-JUL-2002 AGENCOURT 8488559 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6296078 5' , mRNA sequence.

ACCESSION	B0650970	
VERSION	B0650970.1	GI:21775142
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 926)  
NIR-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM2503 row: f column: 15  
High quality sequence stop: 595.

FEATURES	location/Qualifiers
source	1. .926

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone_image="6296078"
/cclone_lib="NIH-MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB1; site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."

```

BASE COUNT	267 a	290 c	218 g	151 t
ORIGIN				
Query Match		31.4%	Score 730.4;	DB 14; Length 926;
Best Local Similarity		93.7%	Pred. No. 3.1e-121;	
Matches 772; Conservative		0; Mismatches 51;	Indels 1;	Gaps 1;

Oy 4 GGGATCACTAGCATGTCGCGAGAGCGGCCCTGGGACGAGATTGAGAAATCTGCCACTA 63

D	b	77	GGGATCTAGCATATCTTCGGGAGGCGGCCCTTGGGACGAGATTGAGAAATCTGCCAGTA	136
O	y	64	ATGGGGGATGGACTAGAAACCTTCCCAATGTCTACACACAGGCCAGGCCCAACCCAG	123
D	b	137	ATGGGGGATGGACTAGAAACCTTCCCAATGTCTACACACAGGCCAGGCCCAACCCAG	196
O	y	124	CCAGGCAACGCAAGCCAGGACCAACCCCGCCCGCCAGAGACTCCAACTTAACAGCC	183
D	b	197	CCAGGCAACGCAAGCCAGGACCAACCCCGCCCGCCAGAGACTCCAACTTAACAGCC	256
O	y	184	AAGAGGCAAGCAACCAACTGCAATACCTGGCTCAGAGTGGTCTCAAGCACTATGSA	243
D	b	257	AAGAGGCAAGCAACCAACTGCAATACCTGGCTCAGAGTGGTCTCAAGCACTATGSA	316
O	y	244	CACGAGTTGCAATGAGCCTTTCCAGCAGCCCTGTGATGGCGTCAAGCTAACCTCCCTGAT	303
D	b	317	CACGAGTTGCAATGAGCCTTTCCAGCAGCCCTGTGATGGCGTCAAGCTAACCTCCCTGAT	376
O	y	304	TACTTAAAGATCATTTAAAACGCTATGGAATATGGAAACAATTAAGAGCGCTTGGAAAC	363
D	b	377	TACTTAAAGATCATTTAAAACGCTATGGAATATGGAAACAATTAAGAGCGCTTGGAAAC	436
O	y	364	AACATTACTGGAATGCTCAGGAATGTATCCAGGACTTCAACACTATGTTTACAATTGT	423
D	b	437	AACATTACTGGAATGCTCAGGAATGTATCCAGGACTTCAACACTATGTTTACAATTGT	496
O	y	424	TACATCTCAACACACCTCGAGATGACATAGTCTTAATGGAGAAAGCTGTGAAAGGTC	483
D	b	497	TACATCTCAACACACCTCGAGATGACATAGTCTTAATGGAGAAAGCTGTGAAAGGTC	556
O	y	484	TTCTTTGCAAAAAATTAATGAGCTTACCACAGAAAGAACCGAGATCATGATAGTCAGCA	543
D	b	557	TTCTTTGCAAAAAATTAATGAGCTTACCACAGAAAGAACCGAGATCATGATAGTCAGCA	616
O	y	544	AAAGGAAGAGGACGTGGGAGGAAAGAAACAGGGACAGCAAAACCTGGCGTTTCCAGGTA	603
D	b	617	AAAGGAAGAGGACGTGGGAGGAAAGAAACAGGGACAGCAAAACCTGGCGTTTCCAGGTA	676
O	y	604	CCAAACACACACTCAAGCATGCACTCTCTCCGAGACCCAGACCCCTCAACCGAATCTCTCT	663
D	b	677	CCAAACACACACTCAAGCATGCACTCTCTCCGAGACCCAGACCCCTCAACCGAATCTCTCT	736
O	y	664	CTGTGCGAGGCCACGCTCACCCCTTCTCCCTGCGCTGACACCCCGGACCTTATGTCAGAGC	723
D	b	737	CTGTGCGAGGCCACGCTCACCCCTTCTCCCTGCGCTGACACCCCGGACCTTATGTCAGAGC	796
O	y	724	CTGTGCTAT - GACAGTGTGCTCTCCCGAGCCACCTGCGAGAGCGCCCGCGCAATGGCCCGCA	782
D	b	797	CTGTGCTATGAGCAGTGTGCTCTCCCGAGCCACCCCATGTGAAAGAGGCCCGCGCGAGAGGCC	856
O	y	783	GCCACAAACCCCGCTCCAGCTCCCGAGCCGCTACAGAGCC	826
D	b	857	CCGAGTCAAAAAACCCCGAGCCGCTTCAAGTTCCTCCCGAGAGGCC	900

RESULT 14	BM983041/c	740 bp	mRNA	linear	EST 21-MAR-2002
LOCUS	BM983041				
DEFINITION	UI-CF-EN1-act-f-11-0-UI-s1 UI-CF-EN1 Homo sapiens cDNA clone				
	UI-CF-EN1-act-f-11-0-UI 3', mRNA sequence.				

ACCESSION	BM983041
VERSION	BM983041.1
KEYWORDS	GI:19607155
SOURCE	EST.
	human.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	
↑ (bases 1 to 740)	

**AUTHORS** Bonaldo, M.F., Lennon, G. and Soares, M.B.  
**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery  
**JOURNAL** Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 FORWARD  
 POLYA-Yes

FEATURES  
 source  
 1. 740  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EN1-act-f-11-0-UI"  
 /clone\_lib="UI-CF-EN1"  
 /tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.  
 TAG\_LIB-UI-CF-EN1  
 TAG\_TISSUE=Human Lung Epithelial Cell lines untreated LPS 6hr to LPS 24h  
 TAG\_SEQ=CTGCTCAGGT"  
 BASE COUNT 115 a 198 c 240 g 184 t 3 others  
 ORIGIN  
 Query Match 31.3%; Score 729; DB 14; Length 740;  
 Best Local Similarity 98.9%; Pval. No. 5.0e-121;  
 Matches 732; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 901 ACCCCACACACATGACACCATTCACGAGCAGCCCTGCGCCCGGAGCCCAAGAC 960  
 |||||||  
 Db 740 ACCCCACACACATGACACCATTCACGAGCAGCCCTGCGCCCGGAGCCCAAGAC 681  
 |||||||  
 QY 961 ACCAAGCTGGCGCAGCGCGGAGAGCAGCGCGCTGAAACCTCCAAAGAGAGCTG 1020  
 |||||||  
 Db 680 ACCAAGCTGGCGCAGCGCGGAGAGCAGCGCGCTGAAACCTCCAAAGAGAGCTG 621  
 |||||||  
 QY 1021 CCCGACTCTAGCAGCAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 |||||||  
 Db 620 CCCGACTCTAGCAGCAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 561  
 |||||||  
 QY 1081 TGCTGAGCGGCATCTCAAGAGAGATGTTTGCCAAAGAGCGCGGCTACAGGCTGGGCC 1140  
 |||||||  
 Db 560 TGCTGAGCGGCATCTCAAGAGAGATGTTTGCCAAAGAGCGCGGCTACAGGCTGGGCC 501  
 |||||||  
 QY 1141 TTCTACAAGCCTGTGAG 1200  
 |||||||  
 Db 500 TTCTACAAGCCTGTGAG 441  
 |||||||  
 QY 1201 CACCCCATGAGACACACATCAAGTCTAACTGAGAGGCCCTGAGTACCGTGATGCT 1260  
 |||||||

Db 440 CACCCCATGAGACACACATCAAGTCTAACTGAGAGGCCCTGAGTACCGTGATGCT 381  
 |||||||  
 QY 1261 CAGAGATTGGTGTGAGAGTCCGATGATGTTCTCAACCTGCTAAAGTACAACTCTCT 1320  
 |||||||  
 Db 380 CAGAGATTGGTGTGAGAGTCCGATGATGTTCTCAACCTGCTAAAGTACAACTCTCT 321  
 |||||||  
 QY 1321 GACCATGAGTGTGAGAGTCCGATGATGTTCTCAACCTGCTAAAGTACAACTCTCT 1380  
 |||||||  
 Db 320 GACCATGAGTGTGAGAGTCCGATGATGTTCTCAACCTGCTAAAGTACAACTCTCT 261  
 |||||||  
 QY 1381 AAGTATGCCGAGCAGCCTGAGAGACCAAGTGTGCTCTCCCGGAGTCCCTCT 1440  
 |||||||  
 Db 260 AAGTATGCCGAGCAGCCTGAGAGACCAAGTGTGCTCTCCCGGAGTCCCTCT 201  
 |||||||  
 QY 1441 CCCACCAAGTGTGAGGCCCGCCCTCATCCAGCAGCAGCAGCAGTATGCTCTGGAC 1500  
 |||||||  
 Db 200 CCCACCAAGTGTGAGGCCCGCCCTCATCCAGCAGCAGCAGTATGCTCTGGAC 141  
 |||||||  
 QY 1501 AGTACAGTTCGACTGATGACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 |||||||  
 Db 140 AGTACAGTTCGACTGATGACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81  
 |||||||  
 QY 1561 GAGAGCTCAAGCCGTCGACGACGACGACGACGACGACGACGACGACGACGACGAC 1620  
 |||||||  
 Db 80 GAGAGCTCAAGCCGTCGACGACGACGACGACGACGACGACGACGACGACGACGAC 21  
 |||||||  
 QY 1621 CCNAAAGAAAAGAGAGAGA 1640  
 |||||||  
 Db 20 CCNAAAGAAAAGAGAGAGA 1

RESULT 15  
 B0654353 938 bp mRNA linear EST 15-JUN-2002  
 LOCUS  
 DEFINITION ACENOUR1-8286504 NIH-MGC-100 Homo sapiens cDNA clone IMAGE:629391  
 5', mRNA sequence.  
 ACCESSION B0654353  
 VERSION B0654353.1 GI:21778525  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 938)  
 NIH-MGC http://mhc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@p5-remail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLCM2512 row: h column: 24  
 High quality sequence stop: 680.

FEATURES  
 source  
 1. 938  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:629391"  
 /clone\_lib="NIH-MGC-100"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Liver; Vector: pT73; Site.1: XhoI; Site.2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NTR\_MGC  
Library.  
BASE COUNT 262 a 301 c 225 g 150 t  
ORIGIN

Query Match 30.7%; Score 715.2; DB 14; Length 938;  
Best Local Similarity 91.6%; Pred. No. 1.7e-118;  
Matches 779; Conservative 0; Mismatches 68; Indels 3; Gaps 2;

QY 4 GGGATCCTAGCATGCTCTGCGAGAGCGGCGCTGGGAGCATGTTGAGAAATCTGCCAGTA 63  
DB 77 GGGATCCTAGCATGCTCTGCGAGAGCGGCGCTGGGAGCATGTTGAGAAATCTGCCAGTA 136  
QY 64 ATGGGGGATGAGCTAGAAAATTCTCCAAATGTCTACAAACAGAGGCCAGGCCCAACCCAG 123  
DB 137 ATGGGGGATGAGCTAGAAAATTCTCCAAATGTCTACAAACAGAGGCCAGGCCCAACCCAG 196  
QY 124 CGAGCCAAAGCGACGACCAAAACCCCGCCCGGAGACCTCCAAACCTTAACAGCC 183  
DB 197 CGAGCCAAAGCGACGACCAAAACCCCGCCCGGAGACCTCCAAACCTTAACAGCC 256  
QY 184 AAGAGGACAGACCAACCACTGCAATCTGCTCAGAGTGTGCTCAAGACATATGAGAA 243  
DB 257 AAGAGGACAGACCAACCACTGCAATCTGCTCAGAGTGTGCTCAAGACATATGAGAA 316  
QY 244 CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGTCAAGCTGAACCTCCCTGAT 303  
DB 317 CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGTCAAGCTGAACCTCCCTGAT 376  
QY 304 TACTATTAAGATCATTAAGAGCGCTATGATGAGAAACAAATAAGAGCGCTTGAGAAAC 363  
DB 377 TACTATTAAGATCATTAAGAGCGCTATGATGAGAAACAAATAAGAGCGCTTGAGAAAC 436  
QY 364 AACTATTACTGAGATGCTCAGAAATGTATCAGAGCTTCAACACTATGTTTACAATGT 423  
DB 437 AACTATTACTGAGATGCTCAGAAATGTATCAGAGCTTCAACACTATGTTTACAATGT 496  
QY 424 TACATCTACAAACAGCTGTGAGATGACATAGTCTTAATGCGAGAGCTCTGAGAAAGCTC 483  
DB 497 TACATCTACAAACAGCTGTGAGATGACATAGTCTTAATGCGAGAGAGCTCTGAGAAAGCTC 556  
QY 484 TTTCTGCAAAAATAATAGATACCCACAGAAACCGAGATCATGATAGTCCAGGCA 543  
DB 557 TTTCTGCAAAAATAATAGATACCCACAGAAACCGAGATCATGATAGTCCAGGCA 616  
QY 544 AAGAGAAAGAGAGCTGGAGAAAGAAACAGGAGCAAGAAACCTGGCTTCCACAGTA 603  
DB 617 AAGAGAAAGAGAGCTGGAGAAAGAAACAGGAGCAAGAAACCTGGCTTCCACAGTA 676  
QY 604 CCAAAACACATCAGATGACTCTCTCGCAGAGACCCAGACCCCTCAGCCGAATCTCTCT 663  
DB 677 CCAAAACACATCAGATGACTCTCTCGCAGAGACCCAGACCCCTCAGCCGAATCTCTCT 736  
QY 664 CCGTGTGAGGCGACGCTACCCCTTCCGCGTACACCCCGGACCTCAT -CGTCCAGAC 722  
DB 737 CCGTGTGAGGCGACGCTACCCCTTCCGCGTACACCCCGGACCTCATCGTCCAGAAC 796  
QY 723 CCTGTGATGA -CAGTGTGCTCCCGCAGCACTGACAGCGCCCGCCAGATGCCCC 780  
DB 797 CCTGTGATGAAGAGTGGGTGCTCCCGCAGCACTGAGAGAGCCCGCCAGATGGCC 856  
QY 781 CAGCCCAAAACCCCGCTCCAGCTCCCGCAGAGCCGTTACAGAGCCACCCACCATCATC 840  
DB 857 CCCCCAGGCGCAAAACCCCGCTCCAGCTCCCGCAGAGCCGTTCCCAAGCCCGGTACAGAGCC 916  
QY 841 GCGGCCAACC 850  
DB 917 CCCCCCCCC 926

Search completed: March 1, 2003, 14:40:45  
Job time : 3262 secs